Investigation of the ESX-4 secretion system interactome of *Mycobacterium tubeculosis*

by
Michelle Smit

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Medical Sciences (Medical Biochemistry) at the University of Stellenbosch

Supervisor: Prof. N. C. Gey van Pittius
Co-supervisor: Prof. R. M. Warren
Faculty of Health Sciences
Department of Biomedical Sciences

December 2010

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2010

Summary

The genome of the pathogen *Mycobacterium tuberculosis* contains five copies of the ESAT-6 (ESX) gene cluster region, which encodes for a novel type VII secretion system. These gene cluster regions, which are directly involved in pathogenicity and phagosomal escape, contain genes encoding exported T-cell antigens ESAT-6 and CFP-10. The mechanism of action of the ESX secretion system however, remains largely unknown. This study focused on ESX gene cluster region 4 (ESX-4), which has been shown to be the most ancestral region and is also present in other species of Mycobacteria and even in other high G+C Gram-positive bacteria, such as *Corynebacterium diptheriae* and *Streptomyces coelicolor*.

This project aimed to investigate the protein-protein interactions of ESX-4 of *M. tuberculosis* in the model organism *Mycobacterium smegmatis* by means of Mycobacterial Protein Fragment Complementation (M-PFC). M-PFC is a two-hybrid technique which employs two cloning vectors, pUAB300 (conferring resistance to hygromycin B) and pUAB400 (conferring resistance to kanamycin). Genes of interest are cloned into these vectors and co-transformed into the model organism *M. smegmatis* after which it is expressed as fusion proteins. Interaction of the proteins allows selective growth on a medium containing the antibiotic trimethoprim. Various interactions were identified throughout this region, including self-interactions as well as the expected interaction between the ESAT-6 and CFP-10 protein family members esxT and esxU. Since this region is ancestral, ESX-4 provides the basic model of the mechanism of secretion of the type VII secretion system. Many similarities were apparent when the interactions identified for ESX-4 were compared to the interactions previously identified in ESX-3.

Interactions identified by means of M-PFC provide a basis for the further study of the structure of this secretion system, and should be confirmed by means of other techniques, such as co-immunoprecipitation. Despite the ability of M-PFC to identify protein-protein interactions in a mycobacterial system, and thus overcoming some of the limitations of the classical yeast two-hybrid model, it must still be regarded as a fishing experiment for potential interactions.

A further aim of the project was to construct a knock-out of ESX-4 in the model organism *M. smegmatis*, which contains three ESX regions, namely ESX-1, -3 and -4. Homologous recombination proved to be an effective technique for the construction of the knock-out, also indicating that ESX-4 is not essential for *in vitro* growth of *M. smegmatis*. The knock-out strain showed no morphological differences to the wild type strain of *M. smegmatis*. The knock-out strain will in future be compared to the wild type strain in various functional studies in order to determine the function of the ancestral ESX region.

Opsomming

Die genoom van die patogeen *Mycobacterium tuberculosis* bavat vyf kopieë van die ESAT-6 geen groep gebiede wat kodeer vir ‘n unieke tipe VII sekresie sisteem. Die geen groep gebiede, wat direk betrokke is by patogenisiteit en fagosomale ontsnapping, bevat gene wat kodeer vir die gesekreteerde T-sel antigene ESAT-6 en CFP-10. Die meganisme van die ESX sekresie sisteem is egter steeds tot ‘n groot mate onbekend. Hierdie studie het gefokus op die ESX geen groep gebied 4 (ESX-4), wat voorheen bepaal is om die vroegste kopie van die gebied te wees en wat ook in ander species van Mikobakterieë en hoë G+C Gram-positiewe bakterieë, soos *Corynebacterium diptheriae* en *Streptomyces coelicolor*, voorkom.

Hierdie projek was daarop gemik om die proteïen-proteïen interaksies van ESX-4 van *M. tuberculosis* in die model organisme *Mycobacterium smegmatis* te ondersoek deur middel van Mikobakteriële Proteïen Fragment Komplementasie (M-PFK). M-PFK is ‘n twee-hibried tegniek wat van twee kloningsvektore, naamlik pUAB300 (wat weerstand teen hygromycin B bied) en pUAB400 (wat weerstand teen kanamycin bied) gebruik maak. Gene van belang word in die vektore ingekloneer en in die model organisme, *M. smegmatis* geko-transformeer, waarna dit as fusieproteïene uitgedruk word. Indien ‘n interaksie tussen die proteïene plaasvind, sal selektiewe groei op ‘n medium wat die antibiotikum trimethoprim bevat, waargeneem word.

Verskeie interaksies is in hierdie gebied geïdentifiseer, insluitende self-interaksies, sowel as die verwagte interaksie tussen die ESAT-6 en CFP-10 proteïen familielede esxT en esxU. Aangesien hierdie gebied die vroegste kopie is, bied ESX-4 die basiese model vir die meganisme van sekresie van die tipe VII sekresie sisteem. Wanneer interaksies wat vir ESX-4 geïdentifiseer is met die wat voorheen vir ESX-3 geïdentifiseer is vergelyk word is daar heelwat ooreenkomste.

Interaksies wat deur middel van M-PFK geïdentifiseer is, verskaf ‘n basis vir die vêrdere studie van interaksies van hierdie gebied, en sal bevestig moet word deur gebruik te maak van aanvullende tegnieke, soos ko-immunopresipitasie. Ten spyte van die vermoë van M-PFK om

proteïen-proteïen interaksies in ‘n mikobakteriële sisteem, wat dus sommige van die beperkings van die klassieke gis twee-hibriedmodel oorkom, te bestudeer, behoort dit steeds as ‘n voorlopige metode van identifikasie beskou te word.

‘n Vêrdere doel van die projek was om ‘n uitslaanmutant van ESX-4 in die model organisme *M. smegmatis*, wat drie van die ESX gebiede, naamlik ESX-1, -3 en -4 bevat, te skep. Homoloë rekombinasie is bewys om ‘n effektiewe tegniek te wees vir die skep van ‘n uitslaanmuntant en het daarop gedui dat ESX-4 nie essensieel is vir die *in vitro* groei van *M. smegmatis* nie. Die uitslaanstam het ook geen morfologiese verskille getoon teenoor die oorspronklike stam nie. Die uitslaanmutant sal in die toekoms gebruik word in ‘n verskeidenheid funksionele studies waar dit vergelyk sal word met die oorspronklike stam, ten einde die funksie van die vroegste ESX-gebied te bepaal.

Acknowledgements

I would like to thank all the people that made it possible for me to pursue this degree.

My sincere gratitude to my supervisor Prof Nico Gey van Pittius and co-supervisor Prof Rob Warren and to all my colleagues in Lab 424 for their encouragement and support.

I would also like to acknowledge the Medical Research Council of South Africa, National Research Foundation of South Africa and the Ernst and Ethel Eriksen Trust for their financial support.

TABLE OF CONTENTS

List of Abbreviations

%	percentage
µg	microgram
µg/ml	micrograms per millilitre
µl	microlitre
µM	micromolar
Amp	ampicillin
BCG	Bacille de Calmette et Guérin
bp	base pair
CFP-10	culture filtrate protein of 10 kDa
DCO	double cross over
dH_2O	distilled water
DNA	deoxyribonucleic acid
E.	*Escherichiae*
ESAT-6	early secreted antigenic target of 6 kDa
EtBr	ethidium bromide
g	gram
g/l	gram per litre
G + C	guanine + cytosine
Hyg	hygromycin B
Kan	kanamycin
KCl	potassium chloride
L	litre
LB	Luria-Bertani medium
mg/ml	milligram per millilitre
M.	*Mycobacterium*
M	molar
mDHFR	murine dihydrofolate reductase
MgCl	magnesium chloride
ml	millilitre

mM	millimolar
M-PFC	Mycobacterial protein fragment complementation
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanogram
nm	nanometre
°C	degree Celsius
OD	optical density
RD	region of difference
rpm	revolutions per minute
SCO	single cross over
TB	tuberculosis
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
ZN	Ziehl-Nielsen

List of figures and tables

Figures

Page Numbers

Tables

CHAPTER ONE

BACKGROUND

1.1 Brief History of Tuberculosis

The pathogen *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), has been present in humans since antiquity. Bone finds provide the earliest evidence for tuberculosis in man and animals, with examples of spinal TB (Pott's disease) dating back to about 8000 BC (Herzog, 1998). The frequency of Egyptian skeletons revealing tubercular deformities are high, which suggests that the disease was common amongst the populations of the time. Similarly, deformed bones have been found in Neolithic sites in Italy and Denmark and countries in the Middle East. This archaeological evidence suggests that TB was prevalent throughout the world approximately 4000 years ago (Smith, 2003).

Thanks to new developments in the molecular analysis of ancient mycobacterial DNA, further advancements have been made in the detection and characterization of *Mycobacterium tuberculosis*. Recently, DNA from five *M. tuberculosis* genetic loci was detected in bone samples from a woman and infant who were buried together in a Neolithic settlement in the Eastern Mediterranean, dating from 9250 – 8160 years ago (Hershkovitz et al., 2008). Furthermore, molecular analysis based on the amplification and verification of the *M. tuberculosis* complex insertion sequence IS*6110* was conducted on bone samples from an ancient Egyptian population. Analysis of samples from Upper Egypt (2120 – 500 BC) and the necropolis of Abydos (3000 BC) suggests that the occurrence of *M. tuberculosis* was relatively frequent in ancient Egypt (Zink et al., 2001).

It is speculated that the first literature mentioning TB was that formulated by the Babylonian monarch Hammurabi between 1948 and 1905 BC and is engraved in cuneiform script on a stone pillar. This text describes a chronic lung disease, which is thought to be TB. More conclusively, the Greek literature of the time of Hippocrates (460 – 370 BC) introduces the concept of phthisis, or consumption (other names for tuberculosis), as the most common disease of the period, combined with a high mortality rate. The opinion of the time was that phthisis was a hereditary, rather than infectious disease, even though Aristotle (384 – 322 BC) believed that it was contagious. Only later Galen (131 – 201 AD) suspected the contagious

nature of phthisis, which then formed part of medical thinking for the next few hundred years (Herzog, 1998).

The 17th century saw the first detailed pathoanatomical description of consumption, with Sylvius de la Boë of Amsterdam (1617 – 1655), being the first to describe the characteristic tubercles as a constant in the lungs and other organs of consumptives (Herzog, 1998).

During the 19th century evidence of infectiousness began to grow. In 1882 Robert Koch (1843 – 1910) produced irrefutable evidence that a specific microbe is the fundamental cause of tuberculosis. Shortly after that, in 1895, another important contribution to the diagnosis of TB was presented in the form of X-rays by Wilhelm Conrad von Röntgen (1845 – 1923). This meant that the presence, development and severity of TB could be accurately monitored and studied for the first time (Herzog, 1998).

1.2 The Symptoms and Treatment of Tuberculosis

Tuberculosis can manifest in many forms, with pulmonary TB accounting for the vast majority of cases. Pulmonary TB has previously been described as consumption and phthisis, both these terms are indicative of severe wasting and the coughing up of blood in later stages of the disease. Spinal tuberculosis, also termed Pott's disease, is marked by spinal deformity, as well as other bone defects. A common manifestation of TB in the Middle Ages was scrofula, or cervical lymphadenitis, which is characterised by the swelling of lymph nodes in the neck. Other forms of extra pulmonary TB include disease of the central nervous system, the urogenital tract, the digestive system and continuously in the form of lupus vulgaris (Smith, 2003).

Treatment of tuberculosis is a relatively new development, considering the presence of the disease throughout the ages. The first treatment for tuberculosis was suggested by Hermann Brehmer in the middle of the 19th century. He suggested that TB, until then believed to be incurable, could be healed by open-air treatment by bringing patients to an "immune place", a

region where there were no known consumptives. With this belief he founded his sanatorium in Göbersdorf in 1854. Many other sanatoria opened later in the 19th century, all of which adopted the strict rest cure regime suggested by Peter Dettweiler in 1893. Even though the benefits of sanatorium treatment, such as the belief that high altitude was essential, can be debated, it did remove infectious patients from their public surroundings, as well as enforce rest and a sensible diet. This regulated lifestyle improved the well-being of patients greatly, however, the long term results were not very encouraging (McCarthy, 2001). Due to these poor long-term results, sanatorium treatment was gradually being supplemented. In the 1930s these supplementary treatments included pneumothorax (as pioneered by Carlo Forlanini) and thoracoplasty. These techniques produced consistently good results, including closure of cavities, conversion to negative sputum and well-preserved lung function (Sakula, 1983).

In the early 1900s two French doctors, Calmette and Guérin were able, by means of over 200 serial passages, to create an attenuated strain of *Mycobacterium bovis*, now known as BCG (bacille Calmette-Guérin). BCG is still in use in children today, although the benefits of vaccination in adults do not produce concise data (Hsu et al., 2003).

The most recent development in the treatment of tuberculosis is drug treatment in the form of antibiotics and antituberculous drugs. Currently treatment for tuberculosis consists of an initial two month phase (Isoniazid, Rifampin, Pyrazinamide, Ethambutol), followed by a four month consolidation phase (Isoniazid and Rifampin) (Jerant et al., 2000). However, drug resistance has become a major factor in the treatment of tuberculosis, with the occurrence of multi-drug resistant tuberculosis (MDR) and extensively drug resistant tuberculosis (XDR) on the rise. As a result, tuberculosis is a major global health burden, causing 1.8 million deaths in 2008 and 9.4 million TB cases in the same year (WHO, 2009). The further risk of HIV/AIDS co-infection is potentially devastating, resulting in a high cost, both financially and in terms of lives.

1.3 The Evolution of the *Mycobacterium tuberculosis* complex

Mycobacterium tuberculosis, the causative agent of tuberculosis, is a member of the *Mycobacterium tuberculosis* complex. The *Mycobacterium tuberculosis* complex consists of a group of species sharing 99.9% similarity at nucleotide level and identical 16S rRNA sequences (Brosch et al., 2002). Despite these similarities, which provide evidence that they all evolved from a common ancestor, they differ greatly in their host organism, phenotype and pathogenicity. *M. tuberculosis*, *M. africanum* and *M. canetti* make up the human pathogens of the complex, while *M. microti* is a rodent pathogen. *M. bovis* can infect a wide variety of mammals, including humans, whereas in the attenuated form (BCG) it is used as a vaccine, which rarely causes disease (Cole, 2002).

It was originally thought that *M. tuberculosis* evolved from *M. bovis*, the agent of bovine tuberculosis, roughly 15 000 years ago, following the domestication of livestock. This new organisation of living in villages, as opposed to a nomadic lifestyle, defined the Neolithic period and it seemed as though the animal pathogen had merely adapted itself to the human host (Cole, 2002). However, whole genome sequencing of *M. tuberculosis* and comparative genomic studies have resulted in a different explanation. *M. bovis* has undergone numerous deletions relative to *M. tuberculosis*, resulting in a smaller genome. This indicates that *M. bovis* is the final member of a lineage separate from *M. tuberculosis*, but with a common progenitor (Brosch et al., 2002).

1.4 Phenotypic and Genomic characteristics of *Mycobacterium tuberculosis*

All mycobacteria can be divided into two groups based on their growth rates. So called fast growers will produce colonies in less than seven days when plated onto solid media, whereas slow growers will only form colonies in more than seven days. In addition, fast growers are usually non-pathogenic and slow growers, including *Mycobacterium tuberculosis*, are usually pathogenic (Prescott et al., 2002).

Mycobacterium tuberculosis is a rod-shaped, Gram-positive organism with a complex cell wall with very high lipid content (Prescott et al., 2002). The surface of the pathogen is also coated with a detergent labile capsule layer, which consists of non-covalently linked glycans, lipids and proteins (Sani et al., 2010).

The complete sequencing of the genome of *Mycobacterium tuberculosis* reference strain H37Rv in 1998 heralded a new era for tuberculosis research. The genome of *M. tuberculosis* H37Rv is approximately 4.4 x 10^6 bp in size and contains approximately 4000 genes (Cole et al., 1998). These 4000 genes account for 91% of the potential coding capacity of the genome and have a characteristically high G+C content of 65.5%. Only around 40% of the genes have known functions and only 16% bear a resemblance to known proteins (Prescott et al., 2002).

Analysis of the sequenced genome of *M. tuberculosis* has revealed several pathogenically important components, which include the ESX gene cluster regions.

1.5 The ESX gene cluster regions

The genome of *M. tuberculosis* contains five copies of the ESX gene cluster region, designated ESX-1 to ESX-5 (Fig 1.1). These gene cluster regions encode for a novel secretion system, namely type VII secretion, which is directly involved in pathogenicity and phagosomal escape (Abdallah et al., 2007). These gene cluster regions contain genes encoding exported T-cell antigens, namely ESAT-6 (early secreted antigenic target of 6kDa) and CFP-10 (culture filtrate protein of 10kDa). Due to this, these regions are often referred to as the ESAT-6 or ESX gene cluster regions (Brodin et al., 2005). It has also been demonstrated that ESAT-6 and CFP-10 form a tight 1:1 complex, which contributes to pathogenicity, since this complex binds to the surface of host cells, thus implying a signalling role (Renshaw et al., 2002).

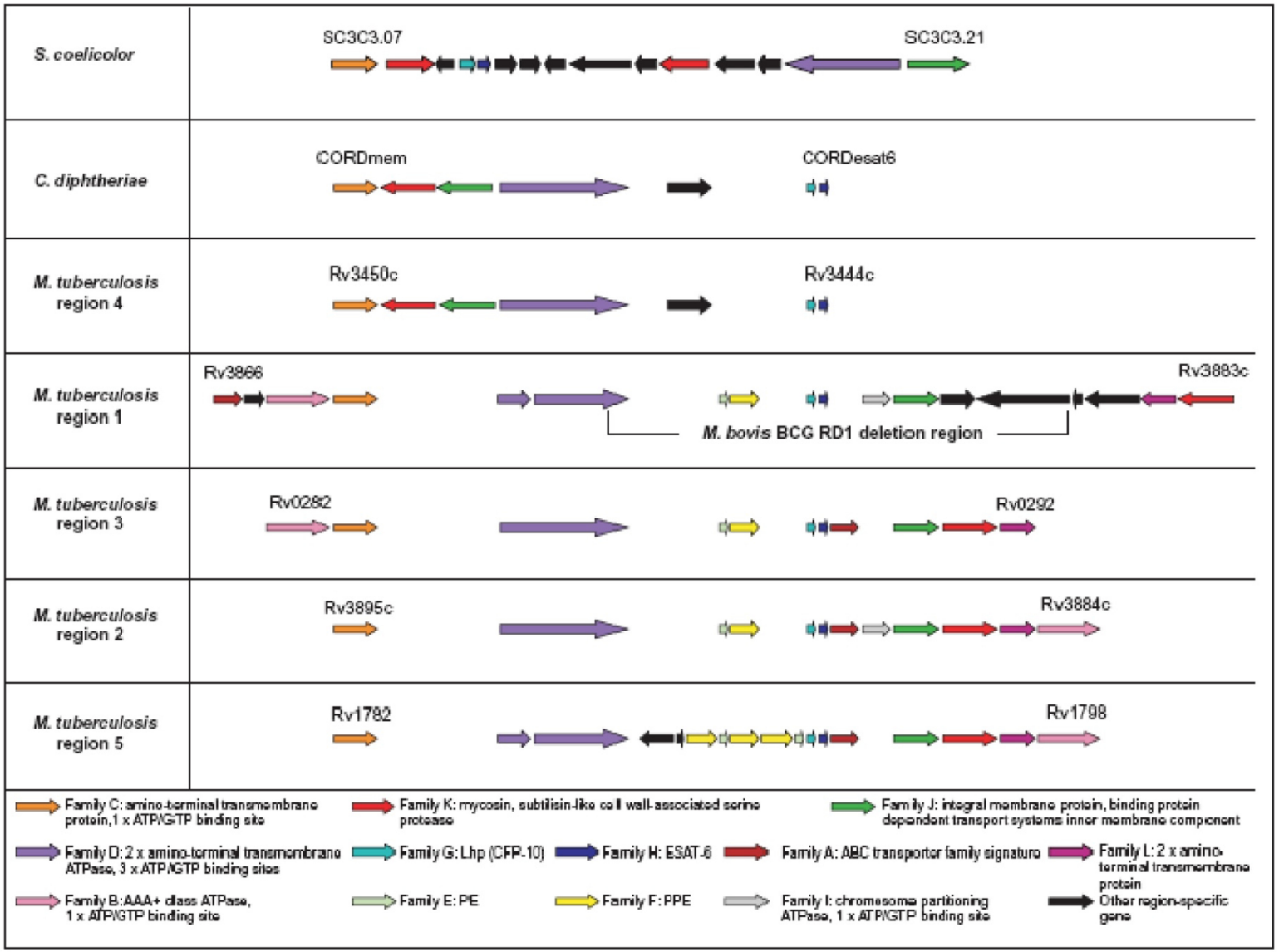
The ESX-1 secretion system is directly implicated in pathogenicity. ESX-1, containing ESAT-6 and CFP-10, is located in the region of difference 1 (RD1) deletion region, which is absent in *Mycobacterium bovis* BCG vaccine strain and results in the attenuation of the organism (Pym et al., 2002).

In addition to ESX-1, the genome of *M. tuberculosis* contains four additional ESX gene clusters, ESX-2 to ESX-5, which are homologous to ESX-1 and also contain genes encoding for ESAT-6 and CFP-10 proteins (Fig 1.1). It has been shown that ESX-4 is the most ancestral of all the regions, present in all other species of Mycobacteria and even in other high G+C Gram-positive bacteria, such as *Corynebacterium diphtheriae* and *Streptomyces coelicolor* (Fig 1.1) (Gey van Pittius et al., 2001).

ESAT-6 region 4 contains seven genes (Rv3444c to Rv3450c). Sequence homology suggest these genes code for proteins with specific functions involved in transport, with Rv3444c (esxT) being an ESAT-6 like protein and Rv3445c (esxU) being a CFP-10 like protein (Fig 1.2). Furthermore, it has been shown that the genes present in ESX-4 are not essential for the *in vivo* survival of *M. tuberculosis* (Singh et al., 2006).

Studying the protein-protein interactions of these regions could provide valuable insights into the structure and function of these secretion systems.

Figure 1.1 Schematic representation of the five ESAT-6 gene cluster regions of *Mycobacterium tuberculosis* in order of duplication. Blocked arrows indicate ORFs, as well as the direction of transcription. The different colours reflect the specific gene family, while the lengths of the arrows reflect the relative gene length. Black arrows indicate unconserved genes present in these regions. Annotations of *M. tuberculosis* H37Rv genes according to Cole et al (1998). The *Mycobacterium bovis* BCG RD1 deletion region is indicated in ESAT-6 region 1. Homology between ESAT-6 region 4 and the regions in *Corynebacterium diphtheriae* and *Streptomyces coelicolor* is also shown. (Gey van Pittius et al., 2001)


S. coelicolor
SC3C3.07
SC3C3.21
C. diphtheriae
CORDmem
CORDesat6
M. tuberculosis region 4
Rv3450c
Rv3444c
M. tuberculosis region 1
Rv3866
Rv3883c
M. bovis BCG RD1 deletion region
M. tuberculosis region 3
Rv0282
Rv0292
M. tuberculosis region 2
Rv3895c
Rv3884c
M. tuberculosis region 5
Rv1782
Rv1798
Family C: amino-terminal transmembrane protein,1 x ATP/GTP binding site
Family K: mycosin, subtilisin-like cell wall-associated serine protease
Family J: integral membrane protein, binding protein dependent transport systems inner membrane component
Family D: 2 x amino-terminal transmembrane ATPase, 3 x ATP/GTP binding sites
Family G: Lhp (CFP-10)
Family H: ESAT-6
Family A: ABC transporter family signature
Family L: 2 x amino-terminal transmembrane protein
Family B: AAA+ class ATPase, 1 x ATP/GTP binding site
Family E: PE
Family F: PPE
Family I: chromosome partitioning ATPase, 1 x ATP/GTP binding site
Other region-specific gene

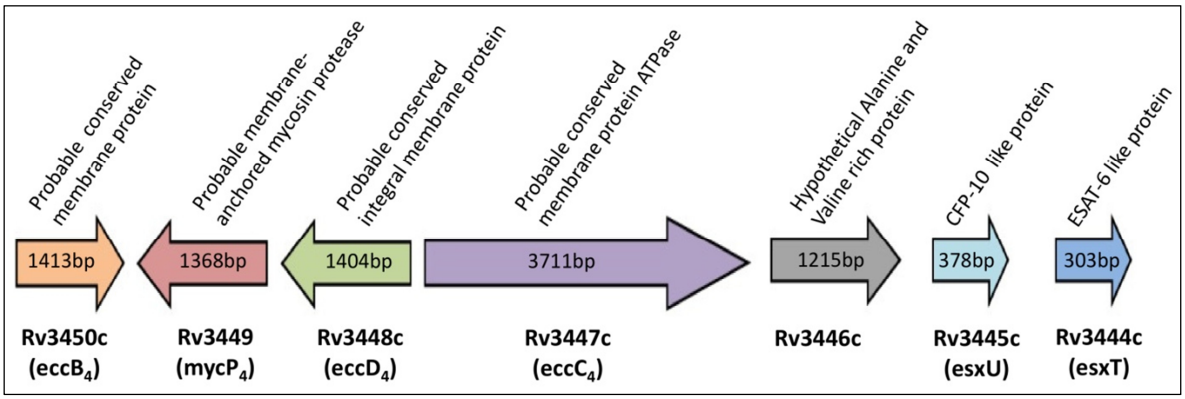


Figure 1.2 A schematic representation of ESX-4 of *Mycobacterium tuberculosis*. The traditional annotation of *M. tuberculosis* H37Rv (Cole et al., 1998) is given in bold and the new nomenclature, as proposed by Bitter et al (2009) is indicated in brackets. Gene length is indicated on the gene and suggested functions are indicated above the genes.

1.6 Studying protein-protein interactions in Mycobacteria

In order to study protein interactions in *M. tuberculosis* Singh et al. (2006) utilised a mycobacterial two-hybrid system, termed Mycobacterial Protein Fragment Complementation (M-PFC). According to this method proteins of interest are independently fused to domains of murine dihydrofolate reductase (mDHFR). Upon interaction of the proteins, functional reconstitution of the two mDHFR domains occurs, resulting in mycobacterial resistance against the antibiotic trimethoprim, thus allowing for selection (Fig 1.3).

To achieve this, genes of interest are cloned into two M-PFC vectors, namely pUAB300 and pUAB400. pUAB300 contains a resistance marker for the antibiotic hygromycin B, whereas pUAB400 contains a resistance marker for the antibiotic kanamycin. These vectors are then co-transformed into *Mycobacterium smegmatis*.

M. smegmatis is a fast-growing, non-pathogenic mycobacterium and also contains a functional ESX protein secretion pathway (Converse and Cox, 2005) (Fig 1.4). This provides an attractive model organism with the advantages of safety to researchers and the ability to study the protein interactions of *M. tuberculosis* in a closely related organism, thus overcoming some of the limitations of traditional two-hybrid systems where yeast is used (Singh et al., 2006).

Using this system, Singh et al. were able to identify the well documented interaction between ESAT-6 and CFP-10, as well as interaction between KdpD and KdpE of *M. tuberculosis*. The system was validated through a *M. tuberculosis* library screen to identify which proteins associated with CFP-10, by which ESAT-6 was again identified (Singh et al., 2006).

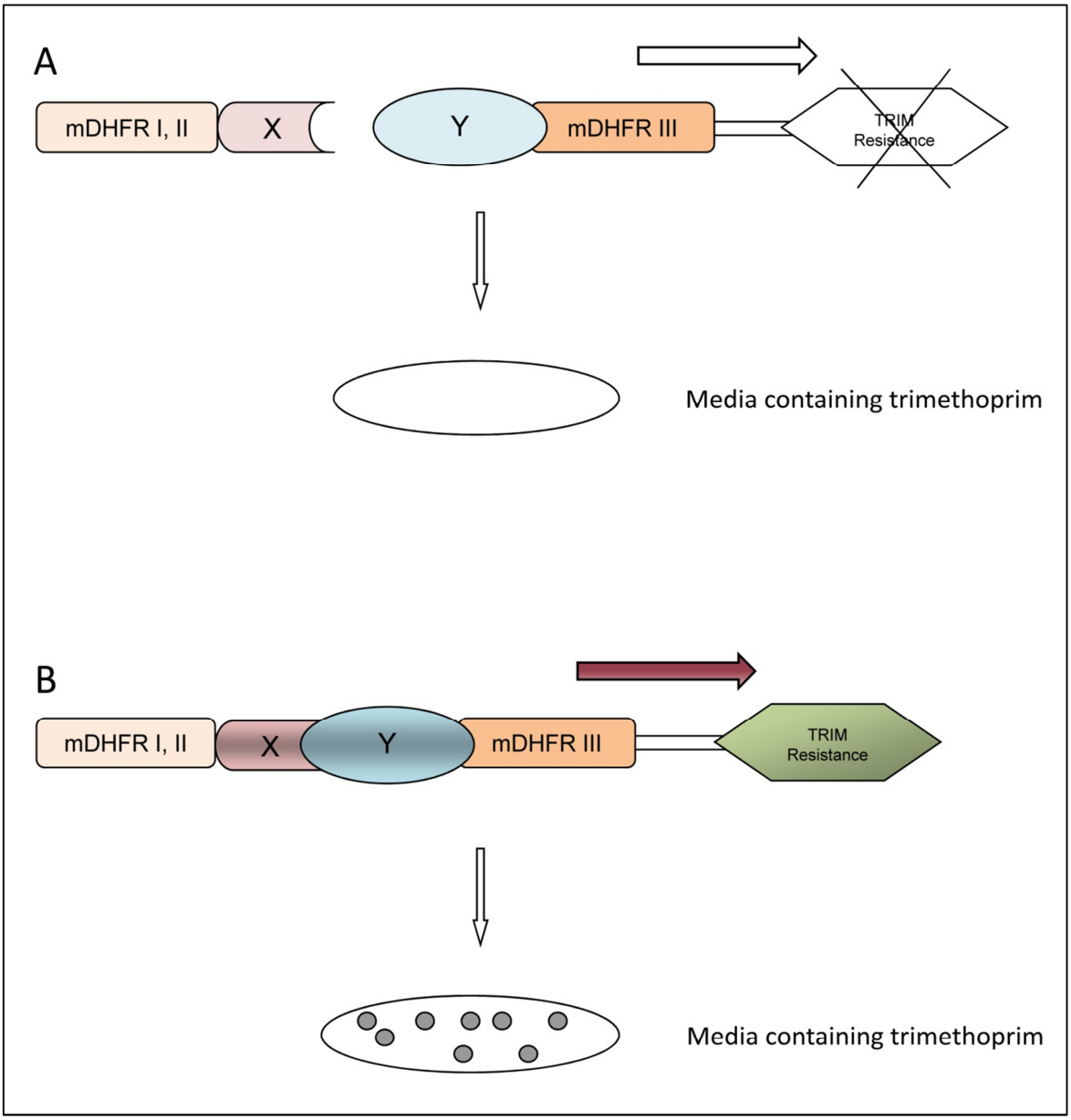


Figure 1.3 Mycobacterial protein fragment complementation (M-PFC) is based on fusing small murine dihydrofolate reductase domains (mDHFR) independently on two possible interacting proteins. If there is no interaction between the proteins (X and Y), growth on plates containing trimethoprim (TRIM) does not occur (a). Upon interaction of the two fused proteins, functional reconstitution of the two mDHFR domains occur and mycobacterial resistance against TRIM will allow growth to be observed (b).

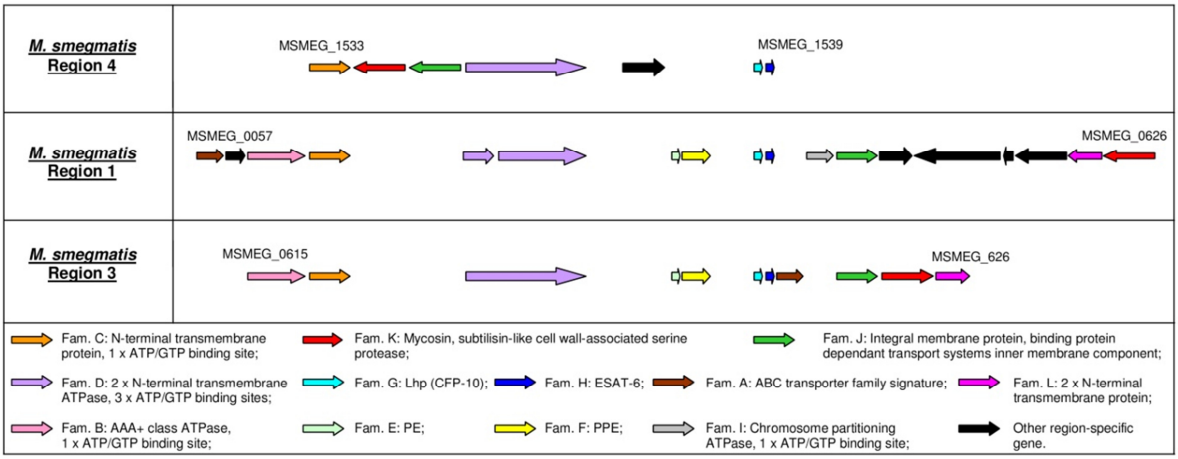


Figure 1.4 The three ESX regions of *M. smegmatis*. Blocked arrows indicate ORFs, as well as the direction of transcription. The different colours reflect the specific gene family, while the lengths of the arrows reflect the relative gene length. Black arrows indicate unconserved genes present in these regions (Gey van Pittius et al., 2001).

1.7 Functional studies in Mycobacteria

In recent years many molecular genetic techniques have been developed to better study the functions of proteins in Mycobacteria. Central to these techniques is the ability to inactivate genes at will, whether it is generating specific, targeted gene knock-outs or random knock-out mutants. These techniques must fulfil the requirements for molecular Koch's postulates, which state that, when testing whether a gene encodes a proposed function, expression of its functional protein must be abolished. This specific function must be restored upon re-introduction of the functional gene. This provides evidence to link a pathogenic phenotype to a causal genotype (Machowski et al., 2005).

The development of techniques to generate targeted gene knock-outs have come a long way since 1990, when Snapper et al., (1990) described a mutant of *M. smegmatis* mc^26 (ATCC607), which presented increased electro transformability. This strain, mc^2155, made

the detection of rare events more likely, which enabled the identification of knock-out mutants in *M. smegmatis*.

The *pyrF* gene was the first gene targeted for knock-out, since, depending on the supplement included in the media – uracil or 5-fluoroorotic acid (5-FOA) – it could be selected either for, or against. *pyrF* mutants could be obtained by transforming a simple suicide plasmid, carrying *pyrF*, and disrupted by a kanamycin resistance gene (*aph*) into *M. smegmatis* (Husson et al., 1990). These initial attempts posed several technical difficulties, which included high frequencies of illegitimate recombination and unstable double crossover (DCO) events (Kalpana et al., 1991, Aldovini et al., 1993).

In 1995 the *ureC* gene of *M. bovis* BCG was successfully inactivated by using a suicide plasmid delivery system. This urease-negative mutant meant that the loss-of-function phenotype could be assayed *in vitro* (Reyrat et al., 1995). In an attempt to reduce the amount of illegitimate recombinations, a system was developed to deliver the mutant allele on a piece of linear DNA 40-50 kb in length. This system was used to generate *leuD* mutant auxotrophs of *M. tuberculosis* (Balasubramanian et al., 1996).

In 1996, Azad et al., managed to replace the *mas* gene of *M. bovis* with the hygromycin cassette. The result was mutants incapable of synthesizing mycocerosic acids. The same method was employed on the *pps* gene cluster of *M. bovis*, resulting in a mutant incapable of synthesizing phthiocerol dimycocerosates (Azad et al., 1997).

In contrast to the techniques developed to create targeted gene knock-outs, random knock-out mutants can be created by random transposon mutagenesis. This requires no prior knowledge of the function of the gene (McAdam et al., 2002).

1.8 Construction of targeted genetic knock-outs

Targeted genetic knock-outs can be constructed by means of allelic exchange based methods. The gene or genetic region to be disrupted is cloned into a suicide plasmid and delivered into cells by means of electroporation. Two recombination events are required to create an allelic exchange mutant. An initial homologous recombination event will result in a single cross over (SCO) recombinant. This recombination between the chromosome and the suicide substrate can occur either upstream or downstream of the mutation and will contain both the wild type and mutant versions of the gene. In addition, all other sequences carried on the vector will be retained.

This is followed by a second recombination event, which will result in a double cross over (DCO) recombinant, which will either produce allelic exchange mutants (knock-outs) or be a reversion to wild type (Machowski et al., 2005).

To better identify allelic exchange mutants, two-step selection / counter-selection procedures have been developed. In 1999, Pavelka and Jacobs refined allelic exchange methodology to produce mutants of two substrains of *M. bovis* BCG, *M. tuberculosis* H37Rv and *M. smegmatis* in which the *lysA* gene had been deleted.

Primary recombinants were selected for on medium containing hygromycin. In addition to the *hyg* gene, the suicide plasmid, pYUB657, also contained a counter selectable marker in the form of the *sacB* gene, which confers sensitivity to sucrose. *sacB* is one of the most commonly used counter selectable markers, specifically in slow growing mycobacteria (Pelicic et al., 1997). Thus, SCOs were not only hygromycin resistant, but also sensitive to sucrose.

In the event of a second homologous recombination event between duplicated regions, the plasmid, (containing the *hyg* and *sacB* genes) is lost. This results in a DCO recombinant that can be either a knock-out or a reversion to wild type. Knock-outs are able to grow on media containing sucrose, but die on media containing hygromycin (Pavelka and Jacobs, 1999).

However, the use of sucrose resistance as a counter selectable marker is hampered by the occurrence of spontaneous *sacB* mutants, which results in false positives being present during counter selection on sucrose plates. To distinguish between spontaneous *sacB* mutants and true knock-outs, the *lacZ* gene can be cloned into the vector. This will allow for the selection of SCOs (which will be blue on a media containing X-gal) and DCOs (which will be white on media containing X-gal) (Parish et al., 1999).

1.9 Problem statement

Relatively little is known about the ESX-4 secretion system of *M. tuberculosis* and other (myco)bacteria. This region contains genes encoding for members of the family of exported T-cell antigens ESAT-6 and CFP-10, which in ESX-1 of *M. tuberculosis* is directly involved in the pathogenicity of the organism. ESX-4 contains the minimal number of genes required for ESX secretion, thus, studying the protein-protein interaction of ESX-4 could help to set up a model of the machinery of this secretion system. Since ESX-4 has also been shown to be the ancestral region, any information generated could provide further insights into the evolution of the ESX secretion systems. In addition, deciphering the original function of the ESX secretion systems will lead to a better understanding of the evolution of the mechanism of pathogenicity.

1.10 Aims of the investigation

This project aims to investigate the protein-protein interactions of the ESAT-6 gene cluster region 4 (ESX-4) of *Mycobacterium tuberculosis* in the model organism *Mycobacterium smegmatis* by means of Mycobacterial Protein Fragment Complementation (M-PFC) (Singh et al., 2006). These interactions can then be used to set up a model for the ESX-4 secretion system, as well as to identify additional substrates of the ESX-4 system.

A further aim of the investigation is to study the function of ESX-4 in *M. smegmatis* by means of comparing wild type *M. smegmatis* to a genetic knock-out of ESX-4 in *M. smegmatis*. A knock-out of ESX-4 in *M. smegmatis* can be constructed by means of homologous recombination techniques (Parish et al., 1999).

It is believed that an understanding of ESX-4 will enable the elucidation of the original function of the gene cluster, as well as to gain insights into the evolution of the ESX secretion system. A better understanding of the mycobacterial secretion system machinery will provide clues as to how to interfere in the action of pathogenicity of this organism.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Bacterial strains and plasmids used in this study

A wide variety of bacterial strains and plasmids were utilised throughout this study. Some bacterial strains and plasmids were commercially available, while others were kind gifts, as indicated in Table 2.1. The cloning vectors pUAB300 and pUAB400 were specific for the Mycobacterial Protein Fragment Complementation protocol and a kind gift from A Steyn. Vectors p2NIL and pGOAL17 were utilised in the construction of a targeted genetic knock-out and a kind gift from E Machowski.

The resistance markers, as well as origins of replication and other features of the vectors utilised in this study are indicated in Fig 2.1

Table 2.1 Bacterial strains and plasmids used in this study.

Strains	**Description**	**Reference**
Bacterial Strains		
E. coli K12 ER2925	Cloning host	New England Biolabs
E. coli JM109	Cloning host	Promega
M. smegmatis mc^2155	Mycobacterial host strain	Snapper *et al*., 1990
M. tuberculosis H37Rv	Laboratory strain of *M. tuberculosis*	ATCC 27294
Plasmids		
pGEM-T Easy	TA cloning vector	Promega
p2NIL	Cloning vector	A kind gift from E Machowski
pGOAL 17	Plasmid carrying *PacI* gene cassette carrying the *LacZ* and *SacB* genes	A kind gift from E Machowski
pUAB300	M-PFC cloning vector	A kind gift from A Steyn
pUAB400	M-PFC cloning vector	A kind gift from A Steyn

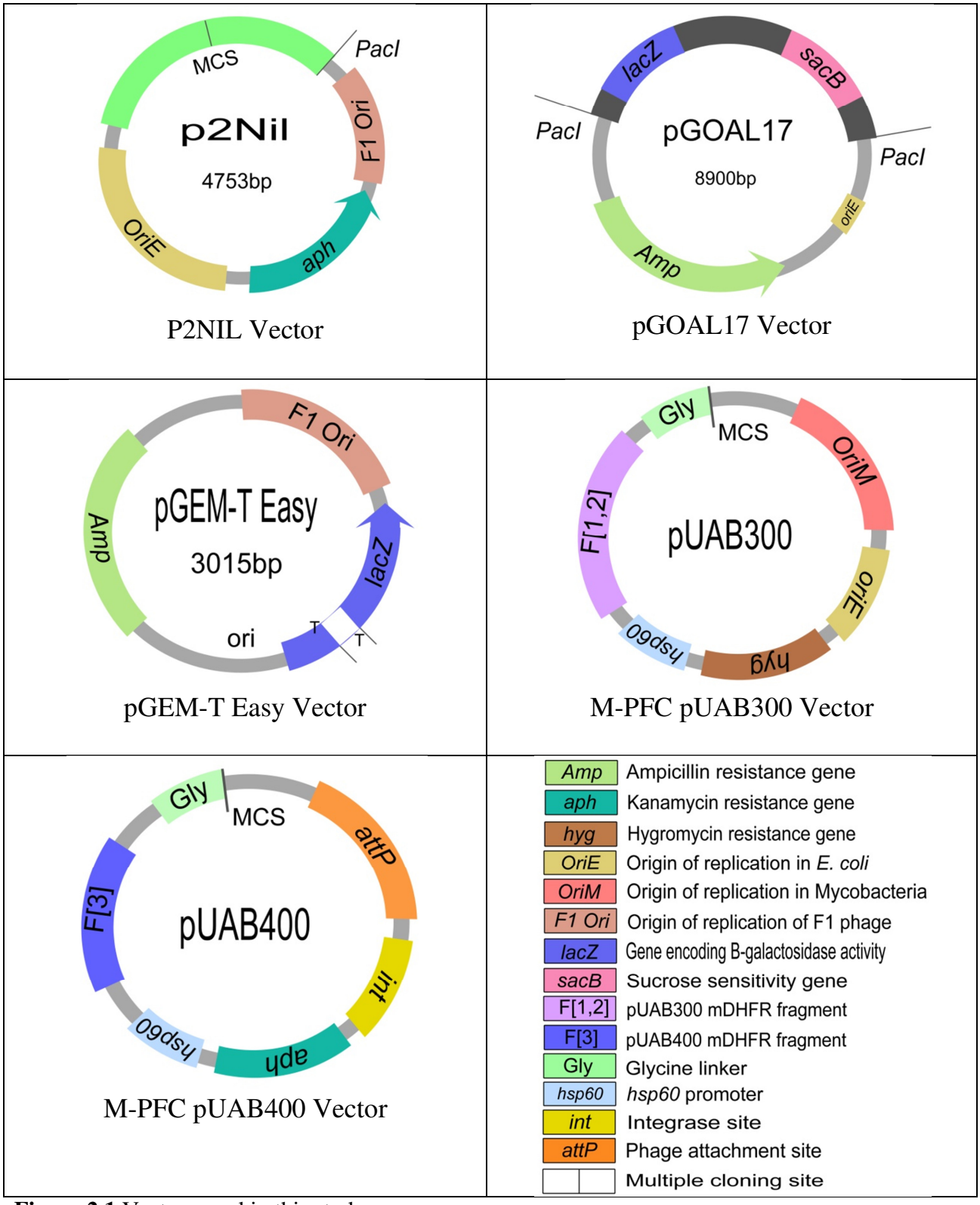


Figure 2.1 Vectors used in this study.

2.2 Mycobacterial Protein Fragment Complementation

2.2.1 PCR amplification and purification

The DNA sequence information for *Mycobacterium* tuberculosis H37Rv was obtained from the publicly available complete genome sequence database at the Pasteur Institute website: http://genolist.pasteur.fr/Tuberculist. Primers were designed for amplification of all 7 genes within ESX-4. All genes had to be amplified with two different forward primers to enable in frame cloning into M-PFC specific vectors, pUAB300 and pUAB400 (given in Table 2.1), respectively. Furthermore, the primers had to provide a cutting site for restriction enzyme *ClaI* (ATCGAT). In addition to amplification of the full genes, both Rv3444c (ESAT-6) and Rv3445c (CFP-10) were also amplified in two parts due to possible toxic effects when the full length is expressed. The sites where Rv3444c and Rv3445c were split are indicated on the modelled 3D structure of each gene, given in Figure 2.2. Since the specific vectors can only efficiently carry an insert of ± 2000 bp, Rv3447c was also amplified in two parts, due to its size (3711 bp). This resulted in a total of 24 constructs. The primer sequences of each construct are given in Table 2.2. For polymerase chain reaction amplification (PCR), an elongation time of 1 minute was used for constructs smaller than 1000 bp and an elongation time of 2 minutes was used for all constructs larger than 1000 bp. FastStart Taq DNA polymerase (Roche) was used for all reactions. Laboratory strain H37Rv DNA was used as a template. Amplification products were run on a 1.5% agarose gel in 1 x SB buffer and visualised under UV light. A molecular weight marker (100 bp; Promega) was loaded on the same gel to confirm that the desired size of the construct had been obtained. All PCR products were purified with the Wizard SV Gel and PCR Cleanup System (Promega).

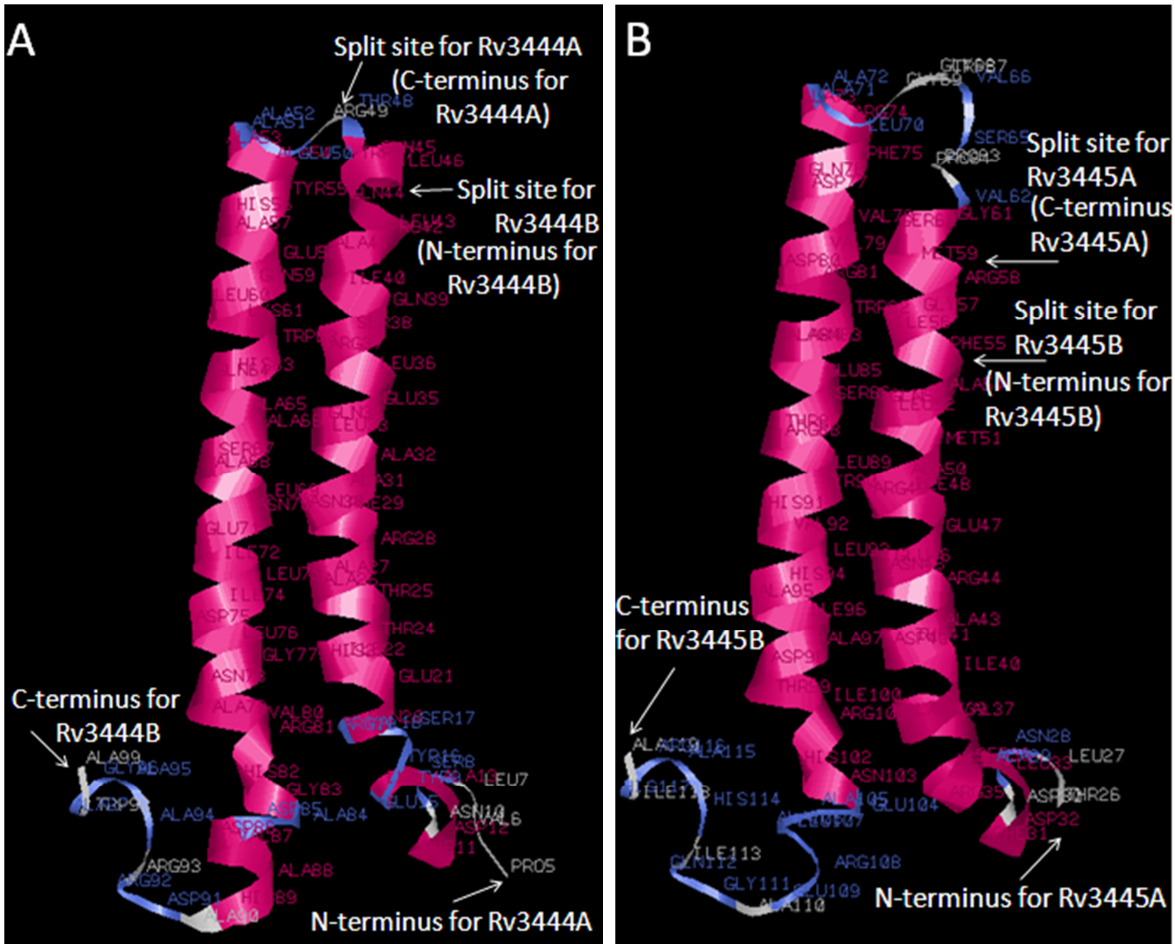


Fig 2.2 Modelled 3D structure of Rv3444 (ESAT-6) (A) and Rv3445 (CFP-10) (B). The sites where Rv3444 and Rv3445 were split in two parts for amplification and the N-terminus and the C-terminus for each part are indicated by white arrows. 3D Modelling was carried out in SWISS-MODEL (http://swissmodel.expasy.org/) (Arnold et al., 2006; Schwede et al., 2003; Guex et al., 1997)

Table 2.2 Primer sequences of each construct according to gene and vector. 'F' denotes the forward primer and 'R' denotes the reverse primer. 'A' and 'B' refers to primers for the amplification of the first and second part of genes split into two parts (Rv3444c, Rv3445c and Rv3447c).

	pUAB300 primers		pUAB400 primers	
Gene	**Primer**	**Primer sequence (5' – 3')**	**Primer**	**Primer sequence (5' – 3')**
Rv3444c	Rv3444pUAB300 FA	ATCGATCTCGACTATGAACGCAGAC	Rv3444pUAB400 FA	ATCGATTCGACTATGAACGCAGAC
	Rv3444pUAB300 FB	ATCGATGCAGCAGCTCTGGACACGG	Rv3444pUAB400 FB	ATCGATCAGCAGCTCTGGACACGG
	Rv3444pUAB300 RA	ATCGATCGTGTCCAGAGCTGCTG	Rv3444pUAB400 RA	ATCGATCGTGTCCAGAGCTGCTG
	Rv3444pUAB300 RB	ATCGATGGAACTAGCGTGCCC	Rv3444pUAB400 RB	ATCGATGGAACTAGCGTGCCC
Rv3445c	Rv3445pUAB300 FA	ATCGATCACGCCGTTGGTTGAACC	Rv3445pUAB400 FA	ATCGATACGCCGTTGGTTGAACC
	Rv3445pUAB300 FB	ATCGATATTCATCGGCCGCATG	Rv3445pUAB400 FB	ATCGATTTCATCGGCCGCATG
	Rv3445pUAB300 RA	ATCGATCTCATGCGGCCGATGAAT	Rv3445pUAB400 RA	ATCGATCTCATGCGGCCGATGAAT
	Rv3445pUAB300 RB	ATCGATGGCTGATCCTCGGTCTATAGG	Rv3445pUAB400 RB	ATCGATGGCTGATCCTCGGTCTATAGG
Rv3446c	Rv3446pUAB300 F	ATCGATCGTGAGCCCACATCGC	Rv3446pUAB400 F	ATCGATGTGAGCCCACATCGC
	Rv3446pUAB300 R	ATCGATCTTCCCGGTTCAACCAAC	Rv3446pUAB400 R	ATCGATCTTCCCGGTTCAACCAAC
Rv3447c	Rv3447pUAB300 FA	ATCGATGCCCAGCATGAATTCAGG	Rv3447pUAB400 FA	ATCGATCCCAGCATGAATTCAGG
	Rv3447pUAB300 FB	ATCGATGAGTCCCTCAGCGGTCCAC	Rv3447pUAB400 FB	ATCGATAGTCCCTCAGCGGTCCAC
	Rv3447pUAB300 RA	ATCGATTGGACCGCTGAGGGACTC	Rv3447pUAB400 RA	ATCGATTGGACCGCTGAGGGACTC
	Rv3447pUAB300 RB	ATCGATCGCGATGTGGGCTCACG	Rv3447pUAB400 RB	ATCGATCGCGATGTGGGCTCACG
Rv3448	Rv3448pUAB300 F	ATCGATGGGGAGAATGCCTACGTCTG	Rv3448pUAB400 F	ATCGATGGGAGAATGCCTACGTCTG
	Rv3448pUAB300 R	ATCGATGGTACGGGACGTGGTCATG	Rv3448pUAB400 R	ATCGATGGTACGGGACGTGGTCATG
Rv3449	Rv3449pUAB300 F	ATCGATCATGACCACGTCCCGTACC	Rv3449pUAB400 F	ATCGATATGACCACGTCCCGTACC
	Rv3449pUAB300 R	ATCGATTAGAGCCAACGCGTCAGTC	Rv3449pUAB400 R	ATCGATTAGAGCCAACGCGTCAGTC
Rv3450c	Rv3450pUAB300 F	ATCGATGCCAGTGCCGAGCCCAG	Rv3450pUAB400 F	ATCGATCCAGTGCCGAGCCCAG
	Rv3450pUAB300 R	ATCGATGCTGATGGCAACCCTGGC	Rv3450pUAB400 R	ATCGATGCTGATGGCAACCCTGGC

2.2.2 Ligation and transformation of *E. coli* cells

All constructs were ligated into the commercial vector pGemT-Easy (Promega). Ligation was allowed to occur overnight at 4ºC. DNA ligations were transformed into electro competent *Escherichia coli* strain K12 ER2925, which is dam negative (dam -), preventing methylation of the *Cla*I restriction site. Transformation of *E. coli* by means of electroporation was carried out in a Bio-Rad Gene Pulser at 2.5 kV, 200 Ohm and 25 µF.

2.2.3 Media and culture conditions

After electroporation *E. coli* K12 ER2925 cells were resuspended in 1ml SOC for an incubation period of one hour at 37ºC with shaking. SOC [SOB (10g tryptone powder (Merck), 5g yeast extract (Fluka Analytical), 5g NaCl (Sigma-Aldrich) and 2.5ml 1M KCl (Merck) dissolved in 900ml distilled water, pH adjusted to 7 with 10M NaOH (Merck) and distilled water added to make up a final volume of 1 litre.) 200µl 1M Glucose and 100µl MgCl was added to 10ml SOB to make SOC]. After one hour of incubation at 37ºC the cells were plated on Luria-Bertani (LB) agar plates [10g tryptone powder (Merck), 5g yeast extract (Fluka Analytical), 5g NaCl (Sigma-Aldrich) and 12g bacteriological agar (Merck) dissolved in 1l distilled water] containing ampicillin (Roche 50µg/ml), and plates were incubated overnight at 37ºC (Fig 2.3).

2.2.4 PCR screening and plasmid purification

PCR screening was performed on single colonies to verify the presence of the desired insert in each construct. Single colonies were picked and allowed to grow overnight at 37ºC with shaking in LB and ampicillin. PCR was carried out with the same set of primers as for the original PCR amplification. Upon confirmation of the constructs, cells were grown overnight at 37ºC with shaking in LB and ampicillin. Plasmid DNA was extracted and purified with the Wizard Plus SV Miniprep DNA Purification System (Promega). The DNA concentration was determined spectrophotometrically using a Nanodrop spectrophotometer (Inqaba Biotech). These concentrations were used to dilute the DNA to a concentration of 100ng/µl, after which samples were sent for sequencing.

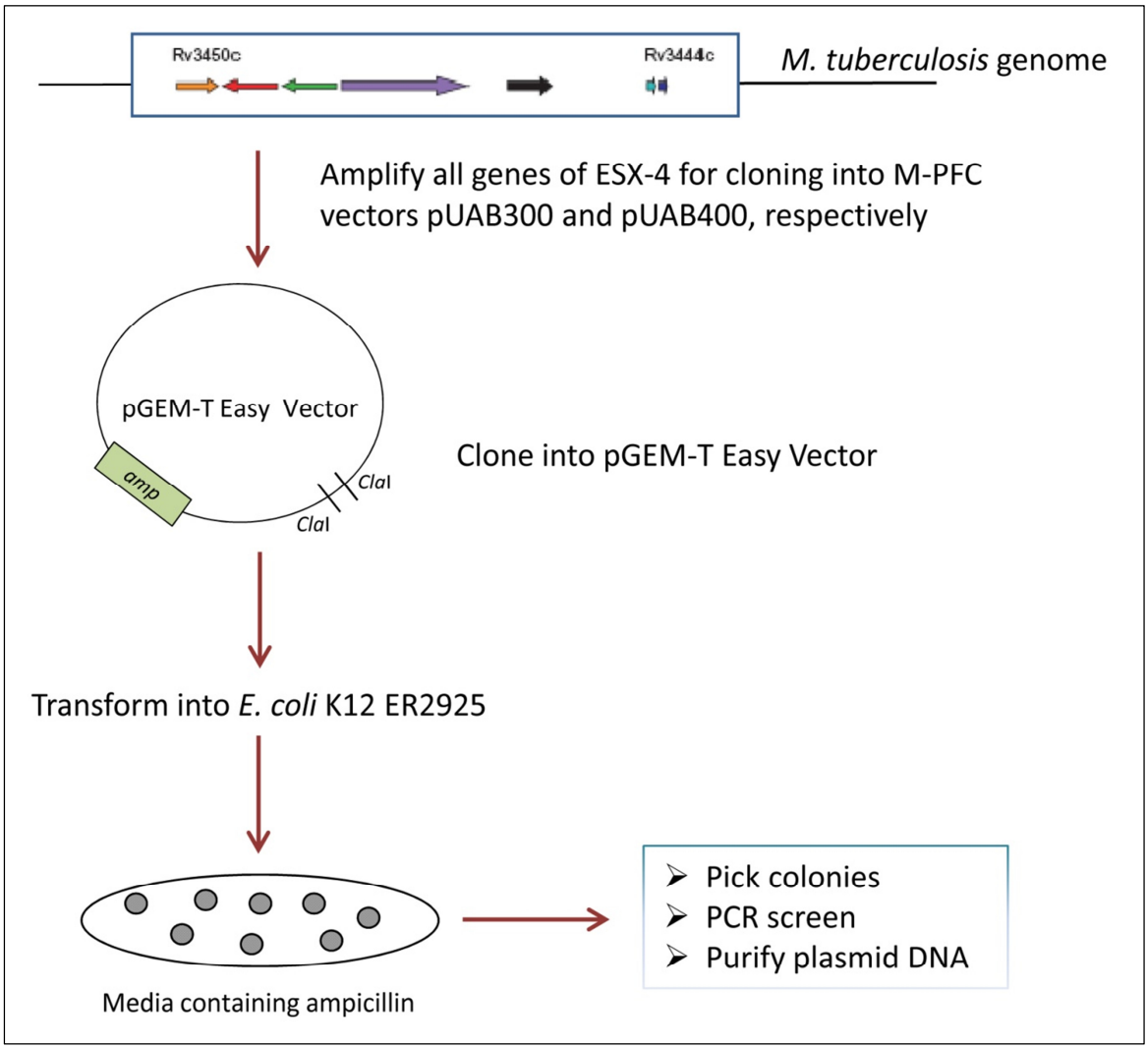


Figure 2.3 Amplification and cloning of genes of ESX-4 into pGEM-T Easy. Subsequent transformation into *E.coli* K12 ER2925 results in growth on medium containing ampicillin.

2.2.5 Preparation of vectors

Cells containing the M-PFC vectors pUAB300 and pUAB400 were grown overnight at 37ºC with shaking. pUAB300 containing cells were grown in LB containing hygromycin B (50µg/ml) and pUAB400 containing cells were grown in LB containing kanamycin (25µg/ml). The plasmids were purified with the NucleoBond Plasmid Purification kit (Macheney-Nagel). The DNA concentration was determined spectrophotometrically using a Nanodrop spectrophotometer (Inqaba Biotech).

2.2.6 *Cla*I digest

All constructs were digested with the restriction enzyme *Cla*I (Promega) at 37ºC for 3 hours to release the inserts. Plasmid vector DNA was digested with *Cla*I at 37ºC for 6 hours. All digests were run on a 1.5% agarose gel and the bands of the desired size were cut out and purified with the Wizard SV Gel and PCR Cleanup System (Promega).

2.2.7 Dephosphorylation of vectors and ligation

The vectors were dephosphorylated with shrimp alkaline phosphatase (Roche) before ligation could take place. Ligation was allowed to occur overnight at 4ºC.

2.2.8 Transformation of M-PFC vectors

DNA ligations were transformed into electro competent *E. coli* strain JM109 at 2.5 kV, 200 Ohm and 25 µF, after which cells were resuspended in 1ml SOC medium and allowed to recover for one hour at 37ºC, with shaking. Cells containing the pUAB300 vectors were plated onto LB agar plates containing hygromycin (50µg/ml) and cells containing the pUAB400 vectors were plated onto LB agar plates containing kanamycin (25µg/ml) (Fig 2.4). Plates were incubated overnight at 37ºC.

2.2.9 PCR screening and plasmid purification

PCR screening was performed to verify presence and orientation of the desired construct. Single colonies were picked and allowed to grow overnight at 37ºC with shaking in LB and hygromycin (pUAB300) or LB and kanamycin (pUAB400). PCR was carried out with the original set of reverse primers for each construct. The original forward primers were

substituted with a sequencing primer for all pUAB300 constructs and a sequencing primer for all pUAB400 constructs. The sequencing primers allowed for amplification of ±80bp of plasmid DNA in addition to the construct, which enabled selection of constructs in the correct orientation. Cells containing constructs in the correct orientation were grown overnight at 37ºC with shaking in LB and hygromycin (pUAB300) or LB and kanamycin (pUAB400). Plasmid DNA was purified with the Wizard Plus SV Miniprep DNA Purification System (Promega). The DNA concentration was determined spectrophotometrically using a Nanodrop spectrophotometer (Inqaba Biotech). These concentrations were used to dilute the DNA to a concentration of 100ng/µl, after which samples were sent for sequencing.

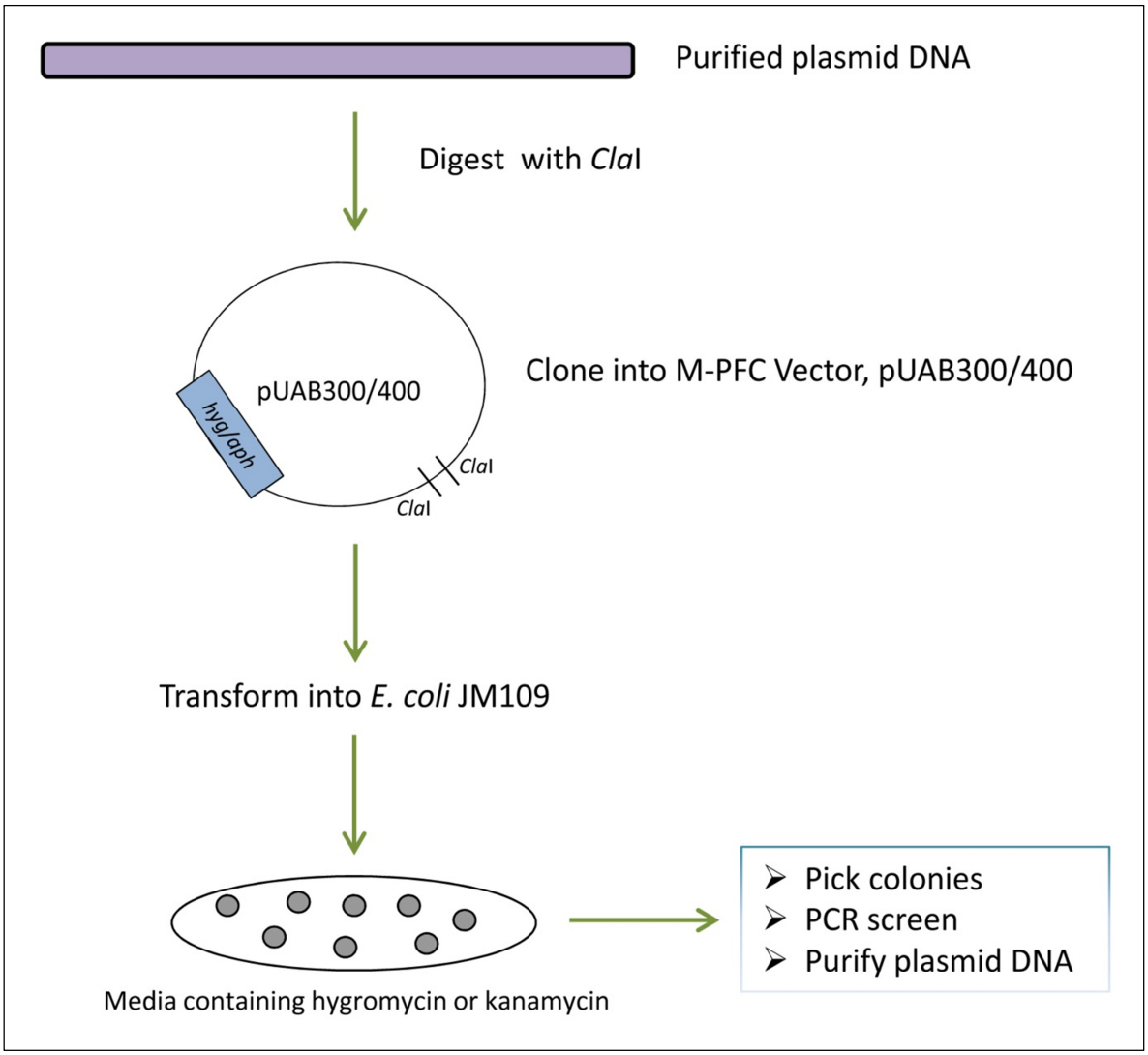


Figure 2.4 *Cla*I digest and cloning into M-PFC vectors, pUAB300 and pUAB400. Subsequent transformation into *E. coli* JM109 results in growth on a medium containing hygromycin (pUAB300) or kanamycin (pUAB400).

2.2.10 Transformation of *M. smegmatis* cells

Transformation of *M. smegmatis* mc^2155 was also carried out by means of electroporation in a Bio-Rad Gene Pulser at 2.5kV, 1000 Ohm and 25µF.

2.2.11 Media and culture conditions

After electroporation *M. smegmatis* mc^2155 cells were resuspended in 1ml 7H9 [2.35g 7H9 (Difco) was dissolved in 485ml distilled water. 5ml Each of 50% glucose (Kimix), 50% glycerol (Merck) and 20% tween-80 (Sigma) were added to cool media] and allowed to recover for three hours at 37°C with shaking. After three hours of incubation cells were plated on 7H11 plates [19g 7H11 (BBL) was added to 970ml distilled water]. 10ml each 50% glucose, 50% glycerol, 20% Tween-80 and appropriate antibiotics (hygromycin, 50µg/ml or kanamycin, 25µg/ml) were added to the media. Plates were incubated at 37ºC for two to three days. Plates were sealed in plastic bags to prevent them from drying out during their incubation period.

2.2.12 Ziehl-Nielsen (ZN) staining

Ziehl-Nielsen (ZN) staining was carried out on all *M. smegmatis* cell cultures prior to transformation to confirm that there is no contamination. A smear of culture was made on a microscope slide and allowed to air dry. The smear was heat fixed. The slide was flooded with carbol fuchsin and heated until plenty of steam came off. After standing for five minutes the slide was washed using tap water and drained. The slide was decoloured with acid-alcohol. After standing for two minutes the slide was washed using tap water and drained. The slide was flooded with the counter stain, methylene blue, and allowed to stand for one to two minutes. The slide was again washed using tap water and drained. The slide was examined under a light microscope after air drying. *Mycobacterium* cells can be identified as pink rods.

2.2.13 Screening of *M. smegmatis* pUAB400/pUAB300 co-transformants

In order to carry out paired screening of pUAB300 and pUAB400 ligated inserts, all pUAB400 (integrating vector) transformations were carried out first (Fig 2.5). PCR screening was performed on single colonies to verify the presence of the pUAB400 ligated vector in *M.*

smegmatis. Single colonies were picked and allowed to grow for two days at 37ºC with shaking in 7H9 liquid medium containing glucose, glycerol, Tween-80 and kanamycin. PCR was carried out with the same set of primers as for the original PCR amplification to confirm the presence of the insert. Upon confirmation of the presence of the vector, cell cultures were supplemented with 10% glycerol and stored at - 80ºC as freeze cultures. ZN staining was carried out on all cultures prior to freezing to confirm absence of contamination.

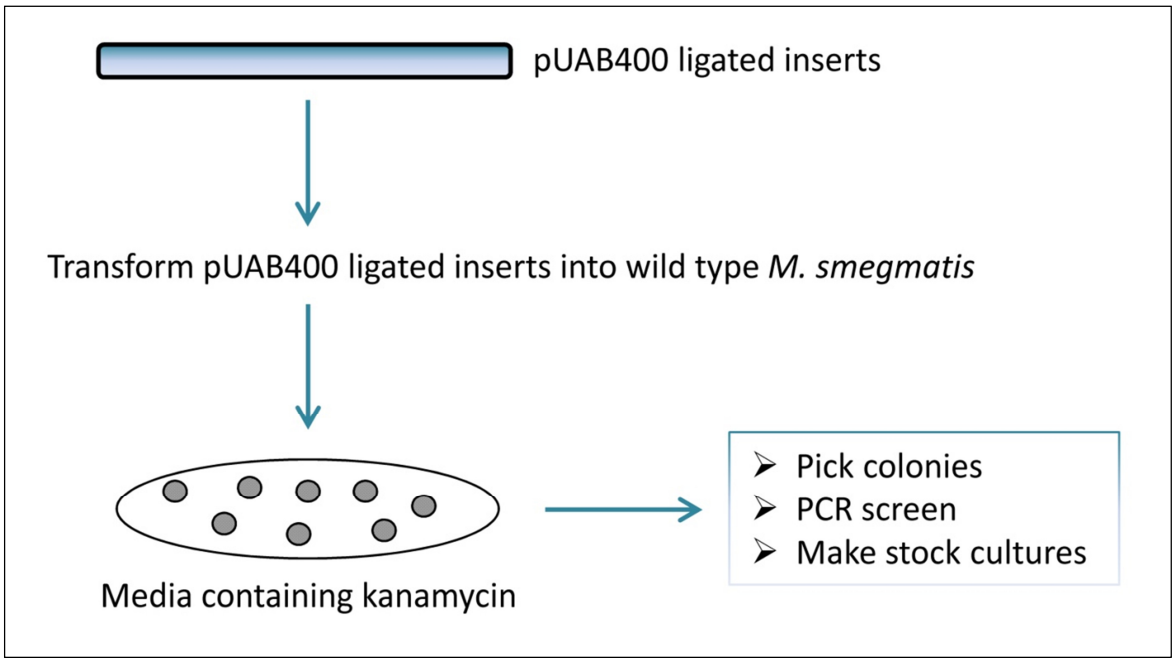


Figure 2.5 Transformation of pUAB400 ligated inserts into wild type *M. smegmatis*, resulting in growth on a medium containing kanamycin.

pUAB400 integrated *M. smegmatis* cultures served as stock for cultures for the subsequent round of pUAB300 transformations. Electro competent pUAB400 *M. smegmatis* cultures were used for co-transformations with pUAB300 insert containing vectors. Co-transformants were allowed to recover at 37°C for three hours and were then plated onto 7H11 plates containing hygromycin, (50µg/ml), kanamycin, (25µg/ml) and trimethoprim, (15µg/ml), and incubated at 37°C for 10 days (Fig 2.6).

Co-transformants producing colonies on 7H11 plates containing hygromycin B, kanamycin, and trimethoprim, were considered as possible interacting partners. Co-transformants containing identical inserts in both M-PFC vectors, which produced colonies on 7H11 plates containing hygromycin B, kanamycin, and trimethoprim were considered as self-interactions. Additional screening of possible interacting partners, as identified by the first screen, was also carried out on 7H11 plates containing an increased concentration of trimethoprim (20µg/ml and 25µg/ml).

Co-transformants on 7H11 plates containing hygromycin, (50µg/ml), kanamycin, (25µg/ml) and trimethoprim, (15µg/ml), were incubated for an additional week (7 days) to determine if any additional growth would occur if plates were incubated beyond the 10 day screening period.

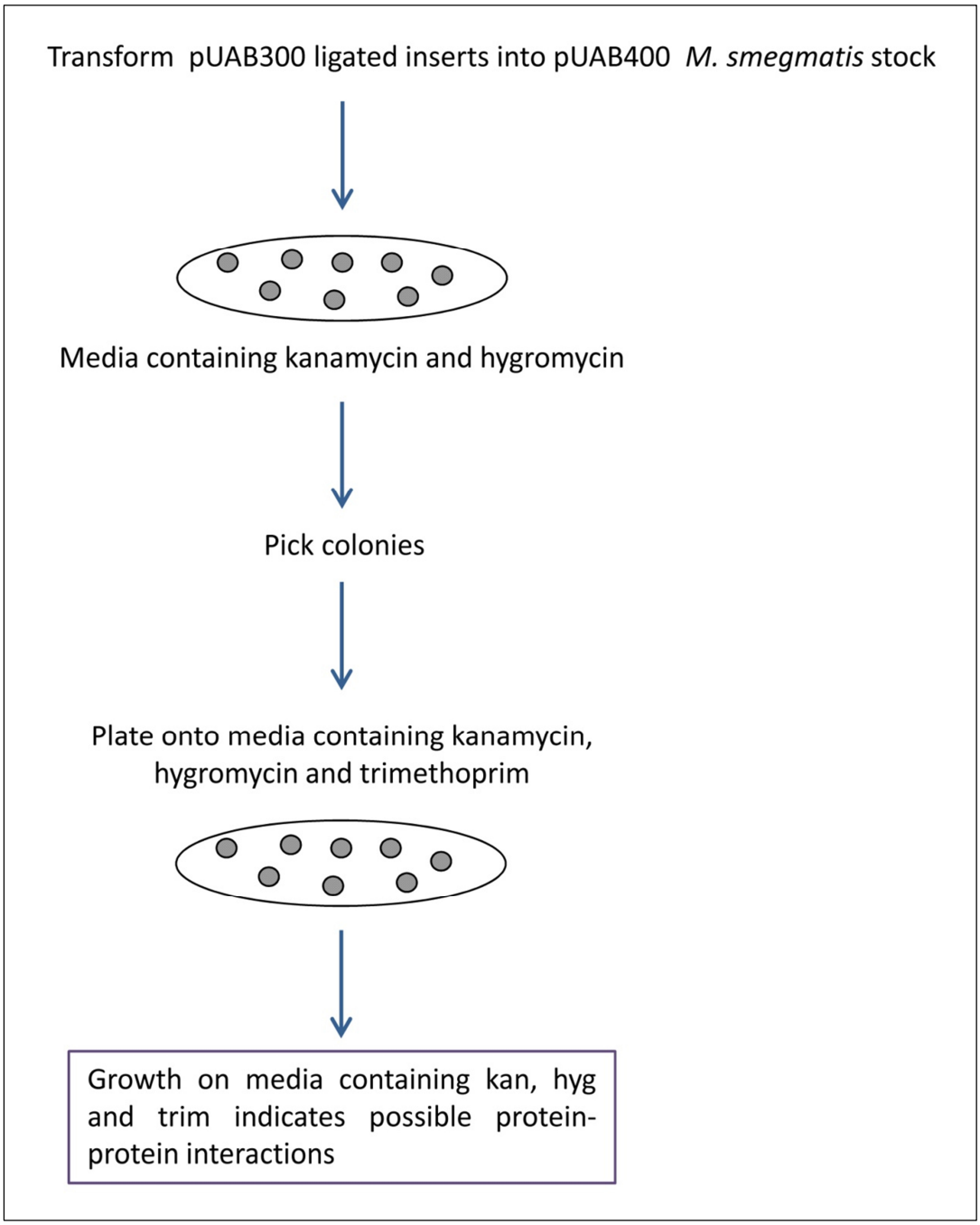


Figure 2.6 Transformation of pUAB300 ligated inserts into pUAB400 *M. smegmatis* stock. Growth on media containing kanamycin and hygromycin indicates the presence of both pUAB400 and pUAB300. Subsequent culturing on media containing kanamycin, hygromycin and trimethoprim indicates possible interacting partners.

2.3 The construction of genetic knock-outs of ESAT-6 Region 4 in *Mycobacterium smegmatis*

2.3.1 Construction of the knock-out construct

The DNA sequence information for *Mycobacterium smegmatis* mc^2155 was obtained from the publicly available complete genome sequence database at The Institute for Genomic Research (TIGR) website: http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=gms.

For the construction of knock-outs of ESAT-6 Region 4 in *Mycobacterium smegmatis* by means of homologous recombination, a genomic region upstream and downstream of ESX-4 in *M. smegmatis*, of approximately 800bp each, was PCR amplified and termed UpstreamMSMEG3450 and DownstreamMSMEG3444. The primers UpstreamMSMEG3450 (forward and reverse) and DownstreamMSMEG3444 (forward and reverse), which contain the restriction sites *Sal*I and *Sfu*I, and *Sfu*I and *Hind*III respectively, were used for amplification (Table 2.3).

The amplified regions were separately cloned into the T vector, pGEM-T Easy. Both the upstream and downstream regions of ESX-4 were confirmed by sequencing. UpstreamMSMEG3450:pGEM-T Easy was digested with *Sal*I (Roche) and *Sfu*I (Roche) and Downstream:pGEM-T Easy was digested with *Sfu*I and *Hind*III (Roche) to release the 800bp insert. The digest was visualised under UV light after gel electrophoresis on a 1% agarose gel. The bands of desired size (≈ 800bp) were cut out and purified with the Wizard SV Gel and PCR Cleanup System (Promega). The purified fragments were subcloned into the corresponding sites of the mycobacterial suicide vector, p2NIL, by means of three way cloning (Fig 2.7).

Table 2.3 Primer sequences used for the construction of the ESAT-6 Region 4 disruption element and screening of SCO and DCO transformants. 'F' denotes the forward primer and 'R' denotes the reverse primer. Restriction sites for knockout construction are highlighted in grey. (GTCGAC being the restriction site for *Sal*I, TTCGAA the restriction site for *Sfu*I and AAGCTT the restriction site for *Hind*III.)

Primer	**Primer sequence (5' – 3')**
UpstreamMSMEG3450-F	GGGGGTCGACGACGAGGAAGAGGCG
UpstreamMSMEG3450-R	GGGGTTCGAACCGTGCTGTGAACGAAACCC
DownstreamMSMEG3444-F	GGGGTTCGAACGGACTGACTCGTTGGAGCG
DownstreamMSMEG3444-R	GGGGAAGCTTCCGCGCAGTCGCCCGT
Region 4 F1-F	GAGATCGCGATGGCCATCGCC
Region 4 F1-R	CGCGTAACCGGTTGCCGTGCC
Region 4 F2-F	CCAACGCCTGGGGCGTCTGAT
Region 4 F2-R	CTGCACGGCAACCTGTCGCAG

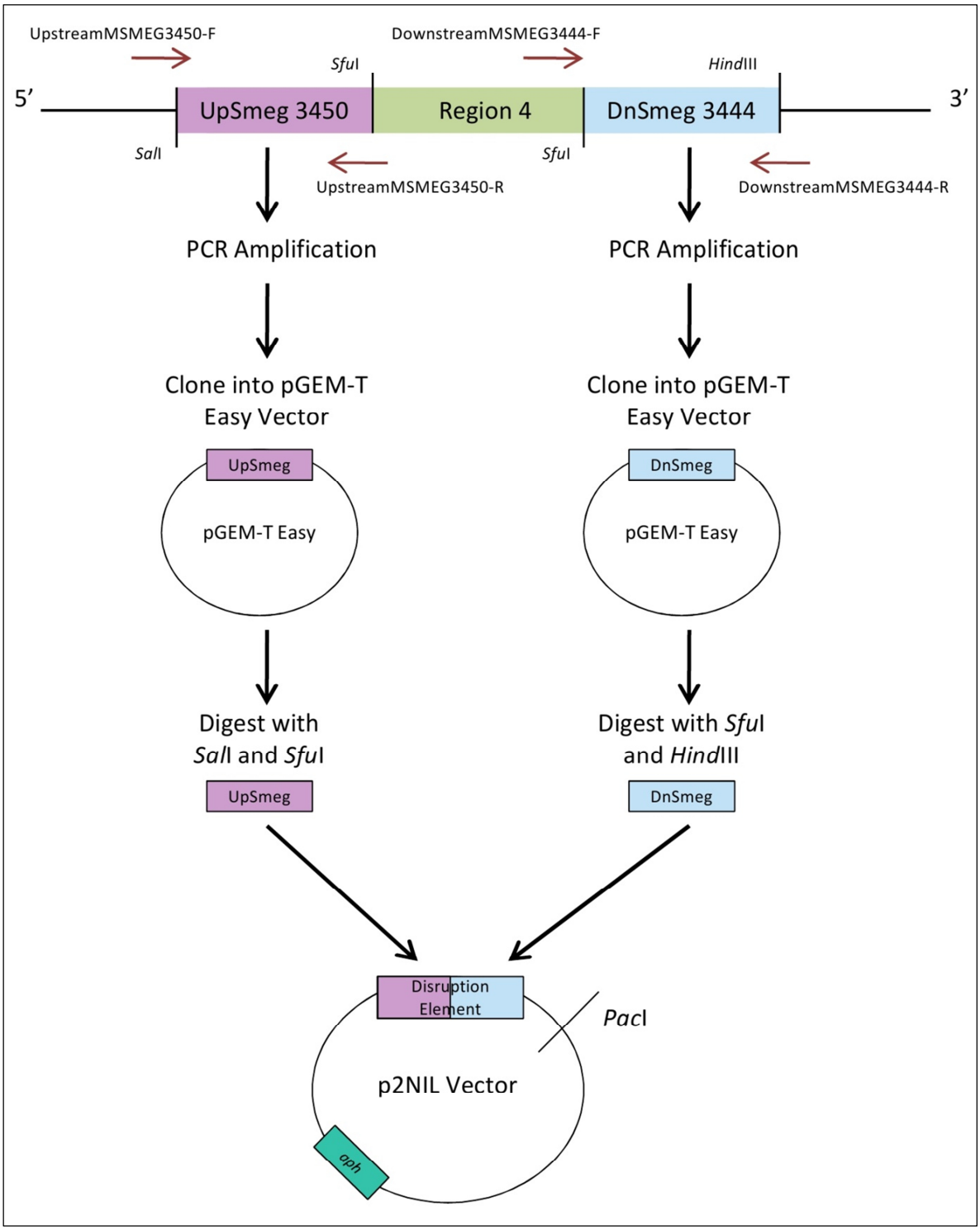


Figure 2.7 Construction of disruption element and subsequent cloning into p2NIL. 'UpSmeg' is indicative of the upstream genomic region of ESX-4 in *M. smegmatis* and 'DnSmeg' is indicative of the downstream genomic region of ESX-4 in *M. smegmatis*.

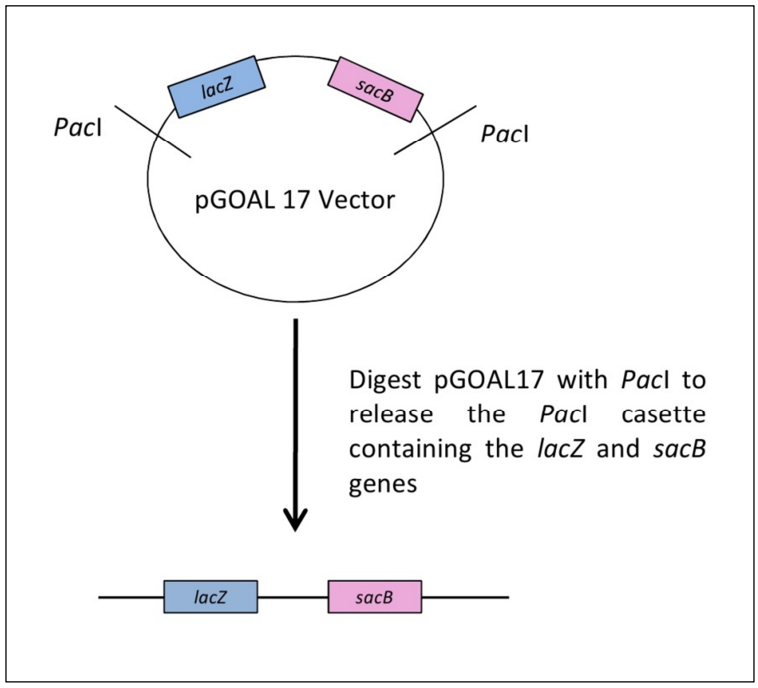


Figure 2.8 Digestion of pGOAL17 releases the *Pac*I cassette containing the *lacZ* and *sacB* genes.

The p2NIL:ΔR4 and the pGOAL 17 vector were both digested with *Pac*I (New England Biolabs). Digestion of pGOAL 17 resulted in two fragments, namely a 2.3kb fragment, the vector band, and a 6.3kb fragment, which is the *Pac*I cassette containing the *lacZ* and s*acB* genes, as described by Parish and Stoker (2000) (Fig 2.8). The resultant *lacZ* gene encodes for β-galactosidase activity, which allows for selection of single cross over (SCO) clones by their blue colour. The s*acB* gene, which encodes for sucrose sensitivity, results in the killing of single cross over intermediate clones, as well as the selection of double cross over (DCO) clones, which will survive on a media containing 5% sucrose (Riedel-de Haën).

The 6.3kb pGOAL17 cassette was cloned into the *Pac*I site for p2NIL:ΔR4 and transformed into electro competent JM109 *E. coli*, after which it was incubated for one hour at 37°C with shaking. The cultures were plated onto LB agar plates, containing 25µg/ml kanamycin (Sigma) and 50µg/ml X-gal (Roche) and incubated overnight at 37°C. Colonies containing the p2NIL:ΔR4:pGOAL17 product was identified as blue colonies on media containing kanamycin and X-gal (Fig 2.9).

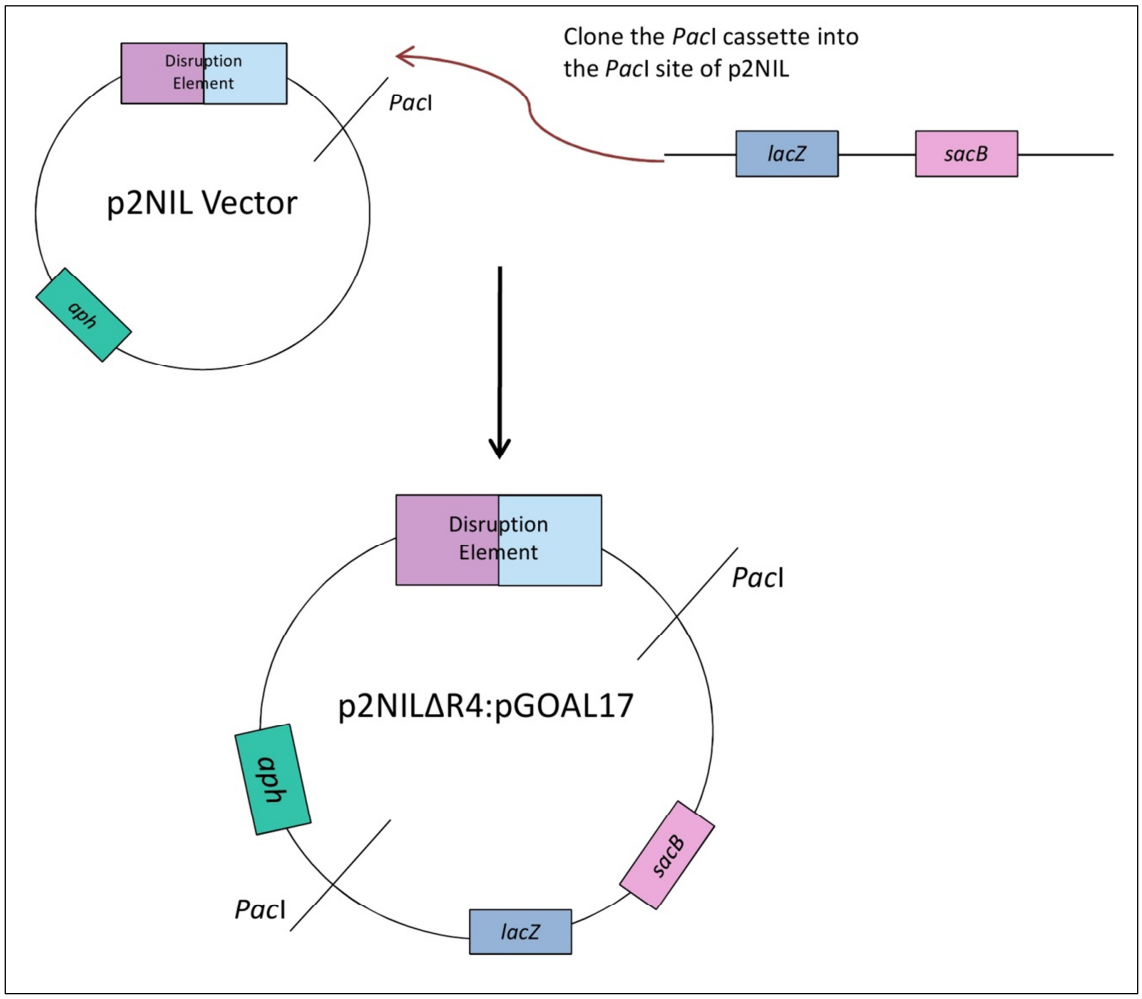


Figure 2.9 The knock-out construct, p2NILΔR4:pGOAL17, containing the disruption element, *aph*, *lacZ* and *sacB* genes.

These colonies were picked and duplicate plated onto LB agar plates containing 5% sucrose to test the sensitivity of the *sacB* gene. Blue colonies, that died on plates containing sucrose were inoculated into 500µl LB and kanamycin (Sigma) and incubated overnight at 37°C with shaking. The presence of the disruption element was confirmed by PCR, resulting in a band of ± 1500bp when visualised under UV light after electrophoresis on a 1% agarose gel. Cultures confirmed to contain the construct were inoculated into 10ml LB, containing kanamycin (Sigma) and incubated overnight at 37°C with shaking. Plasmid DNA was extracted and purified with the Wizard Plus SV Miniprep DNA Purification System (Promega).

2.3.2 Delivery of constructs into *Mycobacterium smegmatis*

Purified constructs were transformed into electro competent *M. smegmatis* mc^2155 and allowed to recover for three hours at 37°C with shaking. Transformants were plated onto LB agar containing 25µg/ml kanamycin (Sigma) and 50µg/ml X-gal and incubated at 37°C for three to four days. A single cross over requires integration of the p2NIL:pGOAL17:ΔR4 construct into *M. smegmatis* and can be identified as blue colonies on LB agar, containing 25µg/ml kanamycin and 50µg/ml X-gal. Blue colonies were picked with sterile pipette tips and inoculated into 500µl LB, containing 25µg/ml kanamycin (Sigma) and 0.1% Tween-80 and inoculated for two to three days at 37°C with shaking. Single cross over transformants were verified by means of PCR (Fig 2.10).

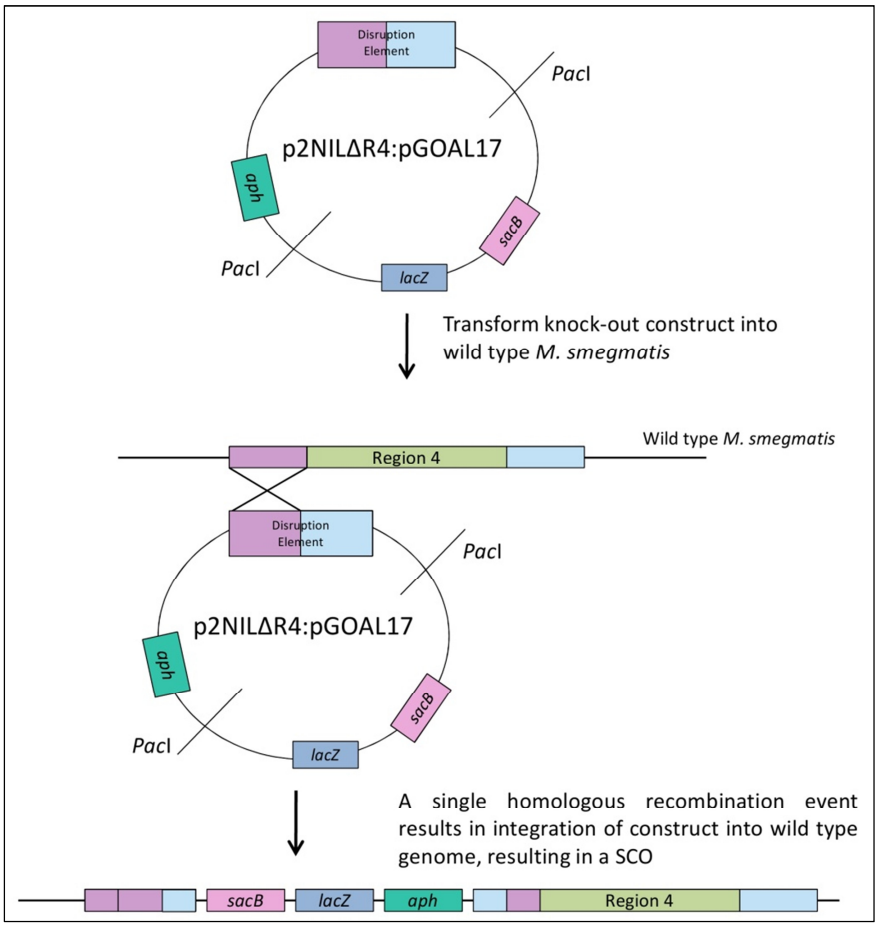


Figure 2.10 Transformation of the knock-out construct into wild type *M. smegmatis*. A single homologous recombination event results in the formation of a single cross over (SCO).

Ten microlitre of confirmed single cross over culture was inoculated into 4ml LB and 0.1% Tween-80, lacking antibiotic to allow for a double cross over event to occur. Inoculants were passaged for two to three days at 37°C with shaking. Ten microlitre of the first passage was transferred into fresh 4ml LB and 0.1% Tween-80, lacking antibiotic and passaged again for two to three days at 37°C with shaking. Serial passaging was continued in this manner to improve chances of inducing a double cross over knock-out event (Fig 2.11).

After two to three days, passaged transformants were plated onto LB agar plates, containing 5% filter sterilized sucrose and 50µg/ml X-gal and incubated for two to three days at 37°C. Double cross over transformants were identified as white colonies on media containing 5% sucrose and X-gal. White colonies were picked with sterile pipette tips and inoculated into 500µl LB + 0.1% Tween-80 and incubated for two to three days at 37°C with shaking. These inoculates were PCR screened to establish whether a double cross over knock-out, or reversion to wild type had occurred. Apparent knock-outs were confirmed by sequencing.

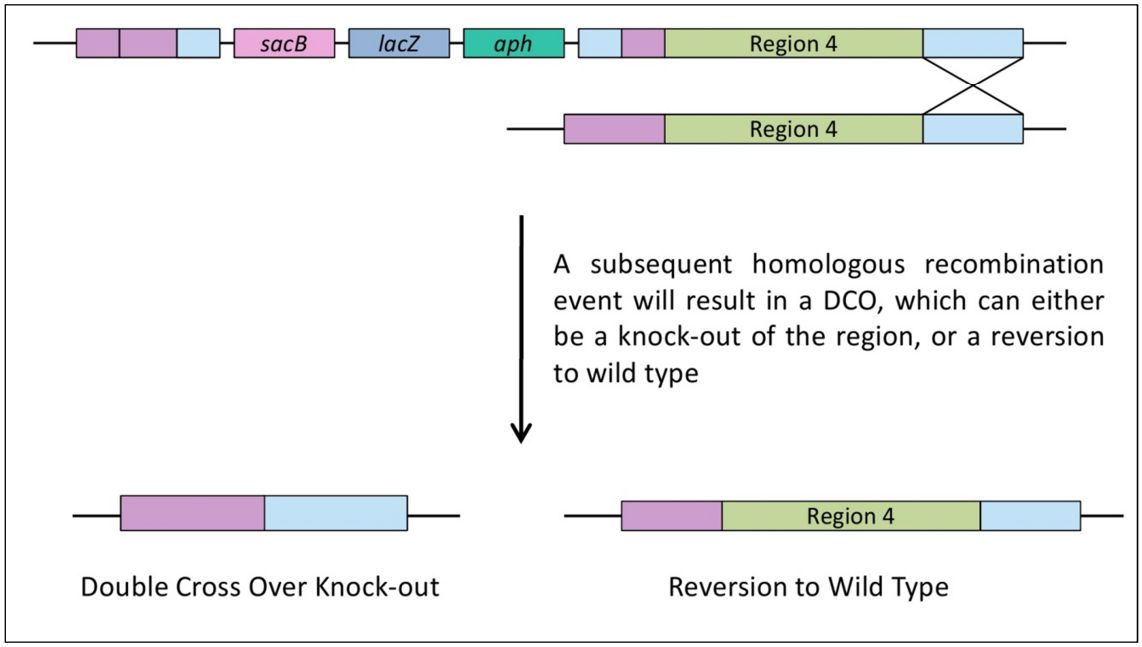


Figure 2.11 Another homologous recombination event will result in the formation of a double cross over (DCO). The DCO can either be a knock-out of the region, or a reversion to wild type.

CHAPTER THREE

RESULTS

3.1 Mycobacterial Protein Fragment Complementation (M-PFC)

3.1.1 PCR amplification and purification

All inserts were amplified from laboratory strain H37Rv genomic DNA with the primer sequences described in Table 2.2. The desired size amplification products were obtained in all cases (Fig 3.1). The PCR products were purified and cloned into the commercial vector pGemT-Easy (Promega).

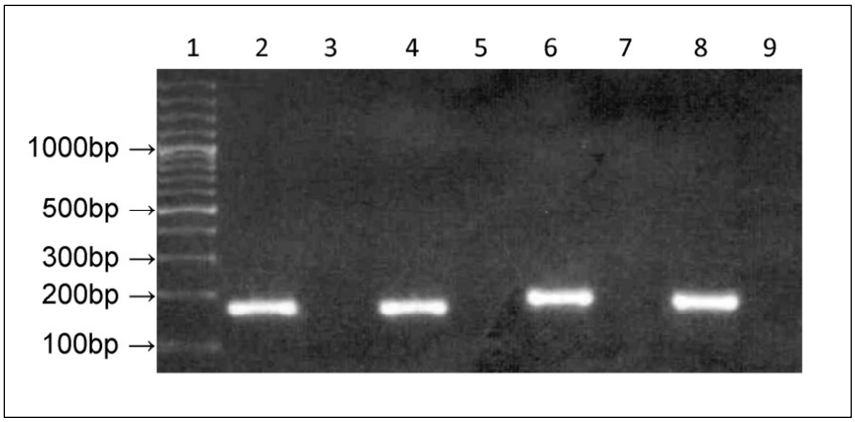


Figure 3.1 An example of amplification of the genes of ESX-4. The PCR products of Rv3444cA were loaded in lanes 2 and 4, while the PCR products of Rv3444cB were loaded in lanes 6 and 8. (Lanes 3, 5, 7 and 9 are no-template controls.)

3.1.2 Transformation, PCR screening and Plasmid purification

Transformations for all constructs were successful, resulting in growth on plates with LB agar and ampicillin. Growth was obtained for all constructs. Colonies were PCR screened with the original primer set (Table 2.2) and colonies containing constructs of the correct size were grown up and the plasmid DNA purified. The determined concentration of plasmid DNA was used to dilute the DNA to 100ng/µl, in a final volume of 20µl and sent for sequencing. Correct sequencing results were obtained for all constructs.

3.1.3 Preparation of vectors, *Cla*I digest and ligation into M-PFC vectors

DNA concentrations of 973.7ng/µl for pUAB300 and 1083.3ng/µl for pUAB400 were obtained from plasmid purification. *Cla*I digestions of all constructs and vectors pUAB300 and pUAB400 resulted in bands of the correct size when run on a 1.5% agarose gel (Fig 3.2). The DNA concentrations obtained after gel cleanup of bands were used to calculate amount of vector, insert DNA and water required for ligation.

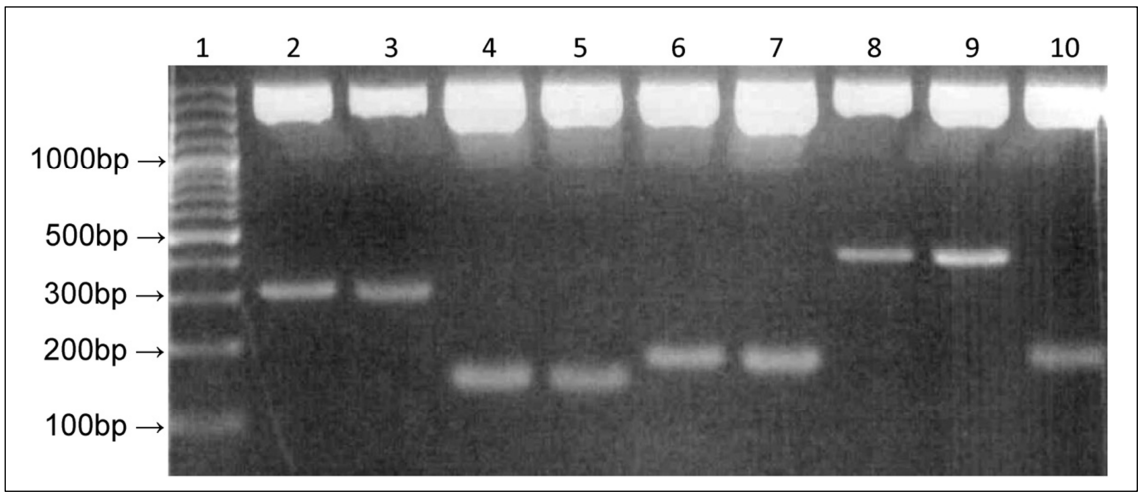


Figure 3.2 *Cla*I digest produced clear separation of bands of desired size. Lanes 2 and 3 contain Rv3444c pUAB300AB and pUAB400AB. Lanes 4 and 5 contain Rv3444c pUAB300A and pUAB400A. Rv3444c pUAB300B and pUAB400B are present in lanes 6 and 7. Rv3445c pUAB300AB and pUAB400AB are shown in lanes 8 and 9, with Rv3445c pUAB300A present in lane 10.

3.1.4 Transformation, PCR screening and Plasmid purification of M-PFC vectors

Transformation of pUAB300 and pUAB400 constructs resulted in growth on LB agar plates containing hygromycin and kanamycin, respectively. PCR screening with pUAB300 and pUAB400 specific primers were used to confirm the presence of the desired insert, as well as to confirm that the insert was in the correct orientation. Colonies containing the correct size constructs were grown up and the plasmid DNA purified. The measured concentration of plasmid DNA was used to dilute the DNA to 100ng/µl, in a final volume of 20µl and sent for sequencing. Sequencing confirmed the identity of the construct's correctness.

3.1.5 Transformation of constructs into *M. smegmatis*

Purified pUAB400 constructs were transformed into wild type *M. smegmatis*, which resulted in growth on 7H11 plates containing 25µg/ml kanamycin. Colonies were PCR screened to confirm the presence of the desired insert and grown up to serve as stock cultures for subsequent transformations of pUAB300 constructs. Transformation of pUAB300 constructs into pUAB400 *M. smegmatis* stock resulted in growth on plates containing 25µg/ml kanamycin and 50µg/ml hygromycin, indicating the presence of both constructs. These colonies were picked and grown in liquid medium containing 25µg/ml kanamycin and 50µg/ml hygromycin. Growth was obtained in liquid culture for all co-transformant colonies, except the ten indicated in Table 3.1.

Table 3.1 pUAB400/pUAB300 co-transformations unable to grow in liquid medium containing 25µg/ml kanamycin and 50µg/ml hygromycin.

pUAB400	pUAB300
3444A	3444
3444A	3444B
3444A	3445B
3444A	3448
3444B	3446
3444B	3450
3445B	3446
3447B	3445
3448	3450
3449	3450

3.1.6 Screening of *M. smegmatis* pUAB400/pUAB300 co-transformants

The 134 pUAB400/pUAB300 co-transformants for which growth was observed in liquid medium containing 25µg/ml kanamycin and 50µg/ml hygromycin were plated onto solid medium containing 25µg/ml kanamycin, 50µg/ml hygromycin and 15µg/ml trimethoprim, and incubated at 37°C for 10 days. Growth on plates containing 15µg/ml trimethoprim is indicated in Table 3.2.

Table 3.2 Potential interacting ESX-4 proteins, identified as growth on media containing 15µg/ml trimethoprim. Potential interacting proteins are indicated by an "X". Pink shading indicates self-interaction and green shading indicates growth for genes in both vector combinations. Grey shading indicates combinations for which growth in liquid culture was not observed.

		pUAB300 Construct											
		3444	3444A	3444B	3445	3445A	3445B	3446	3447A	3447B	3448	3449	3450
pUAB400 Construct	3444	X	X	X	X	X	X		X				
	3444A					X							
	3444B												
	3445	X			X								
	3445A			X						X			
	3445B	X				X	X		X	X		X	X
	3446			X				X	X	X	X	X	
	3447A		X		X	X			X		X		
	3447B	X	X	X		X	X	X	X	X	X	X	X
	3448	X		X	X		X			X	X	X	
	3449	X	X	X	X	X	X		X		X		
	3450	X	X	X	X	X	X	X	X	X	X	X	X

3.1.7 Further screening of *M. smegmatis* pUAB400/pUAB300 co-transformants

The 134 pUAB400/pUAB300 co-transformants that showed growth in liquid medium containing 25µg/ml kanamycin and 50µg/ml hygromycin were plated onto solid medium containing 25µg/ml kanamycin, 50µg/ml hygromycin and 20µg/ml trimethoprim, and incubated at 37°C for 10 days. Growth on plates containing 20µg/ml trimethoprim is indicated in Table 3.3.

The pUAB400/pUAB300 co-transformants were also plated onto solid medium containing 25µg/ml kanamycin, 50µg/ml hygromycin and 25µg/ml trimethoprim, and incubated at 37°C for 10 days, however, extremely little growth was observed when the concentration of trimethoprim was increased to 25µg/ml.

When incubated for an additional week (7 days), no additional growth was observed for co-transformants plated solid on medium containing 25µg/ml kanamycin, 50µg/ml hygromycin and 15µg/ml trimethoprim.

Table 3.3 Potential interacting ESX-4 proteins, identified as growth on a medium containing 20µg/ml trimethoprim. Potential interacting proteins are indicated by an "X". Pink shading indicates self-interaction and green shading indicates growth for genes in both vector combinations. Grey shading indicates combinations for which growth in liquid culture could not be achieved.

		pUAB300 Construct											
		3444	3444A	3444B	3445	3445A	3445B	3446	3447A	3447B	3448	3449	3450
pUAB400 Construct	3444	X		X					X		X		
	3444A				X								
	3444B												
	3445	X											
	3445A												
	3445B												
	3446												X
	3447A								X				
	3447B		X	X		X	X	X	X	X	X		
	3448	X			X								
	3449												
	3450	X	X	X		X	X		X	X		X	

3.2 The construction of genetic knock-outs of ESAT-6 Region 4 in *Mycobacterium smegmatis*

The primers listed in Table 2.3 were used to PCR amplify ± 800bp of the upstream and downstream regions flanking ESAT-6 Region 4 in *Mycobacterium smegmatis*. Genomic DNA of *M. smegmatis* served as template for these amplifications and PCR resulted in a ± 800bp amplification product for the upstream region, termed UpstreamMSMEG3450 and a ± 800bp product for the downstream region, termed DownstreamMSMEG3444.

The amplified regions were cloned separately into the T vector, pGEM-T Easy. The products were confirmed by sequencing. The purified products were subcloned into the corresponding sites of p2NIL, the mycobacterial suicide vector, by means of three way cloning. This resulted in a p2NIL:ΔR4, which was subsequently digested with *Pac*I (Fig B.3). The pGOAL17 vector was also digested with *Pac*I. The resultant pGOAL17 cassette was cloned into p2NIL: ΔR4, resulting in the p2NIL:pGOAL17: ΔR4 construct.

The p2NIL:pGOAL17: ΔR4 construct was transformed into electro competent *E. coli*. JM109. Colonies containing the p2NIL:pGOAL17: ΔR4 construct were identified by their blue colour on LB agar plates containing 25µg/ml kanamycin and 50µg/ml X-gal. These colonies were duplicate plated on LB agar plates containing 5% sucrose. Blue colonies that showed sucrose sensitivity were inoculated into LB containing kanamycin. The presence of the disruption element was confirmed by PCR and resulted in a product of ± 1400bp. The p2NIL:pGOAL17:ΔR4 plasmid was extracted and purified with the Wizard Plus SV Miniprep DNA Purification System (Promega).

The purified constructs were transformed into electro competent *M. smegmatis* by means of electroporation. Single cross over transformants were identified as blue colonies on LB agar containing 25µg/ml kanamycin and 50µg/ml X-gal and verified by means of PCR (Fig 3.3 and Fig 3.4).

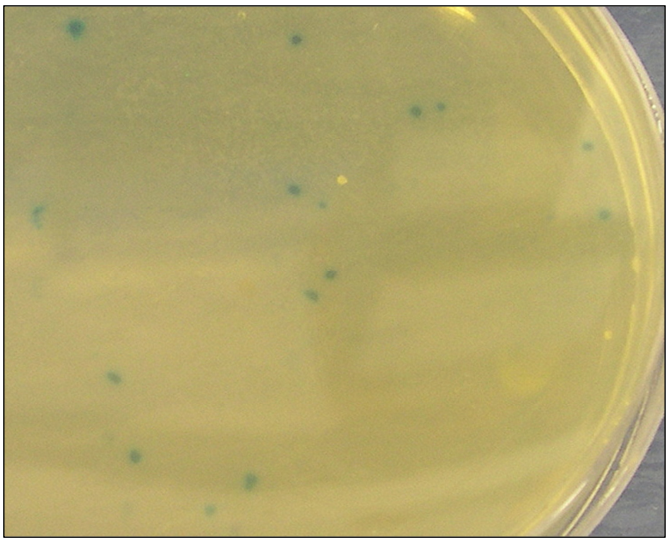


Figure 3.3 Growth of transformants on LB agar containing 25µg/ml kanamycin and 50µg/ml X-gal. Blue colonies represent single cross over mutants.

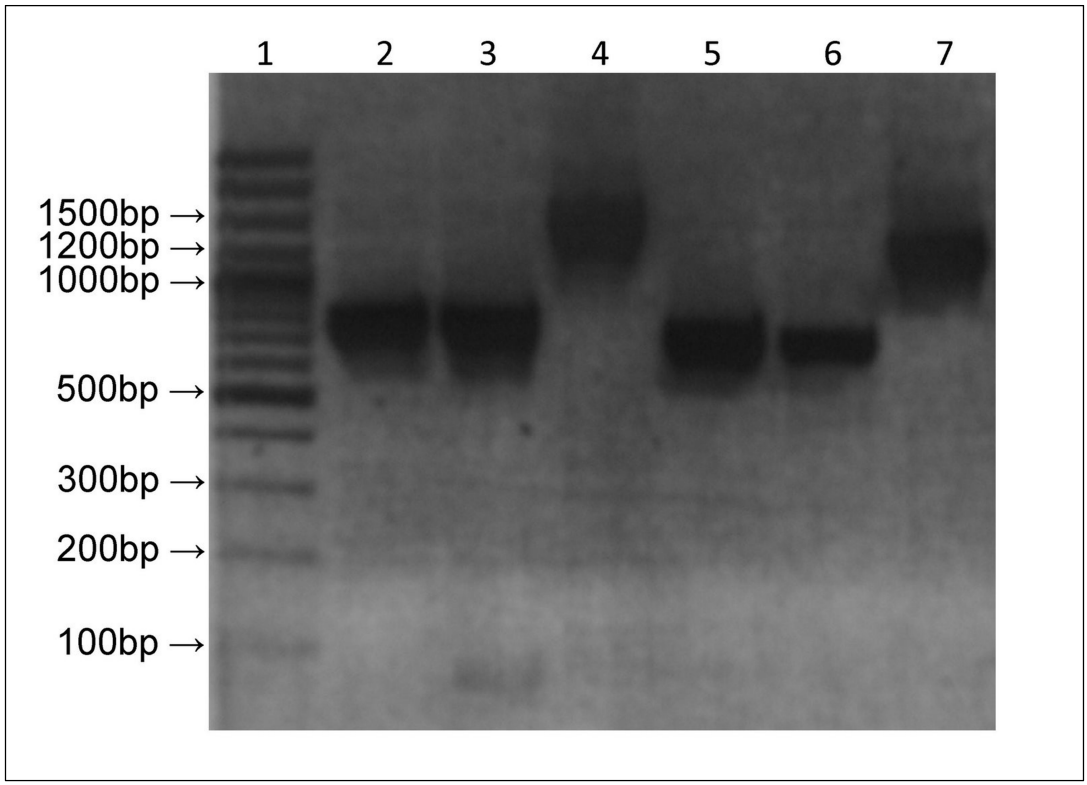


Figure 3.4 Confirmation of SCO of PCR screening. Lanes 2 and 3 contain the ± 800bp products of the original Upstream and Downstream primers for Region 4 (Table 2.3), indicating the presence of the upstream and the downstream regions, respectively. Lane 4 contains the product of PCR amplification with the Upstream forward primer and the Downstream reverse primer of Region 4 (Table 2.3), resulting in a product of ± 1600bp. Lanes 5-7 contain the products of PCR amplification with region 4 specific F1 and F2 primers (Table 2.3). The product sizes of ± 800bp (lanes 5 and 6) and ± 1400bp confirm the SCO.

These single cross over mutants were passaged in LB and 0.1% Tween-80, lacking antibiotic, to allow for a double cross over event to occur. Double cross over transformants were identified as white colonies on LB agar containing 5% filter sterilized sucrose and 50µg/ml X-gal (Fig 3.5). Apparent double cross over mutants were PCR screened to establish whether a double cross over knock-out or a reversion to wild type had occurred (Fig 3.6). Double cross over knock-outs produced a product of ± 1400bp when screened with the UpstreamMSMEG3450 forward and DownstreamMSMEG3444 reverse primer, but no products for the original UpstreamMSMEG3450 (forward and reverse) and DownstreamMSMEG3444 (forward and reverse) primer sets (Fig 2.3). These double cross over knock outs were verified by means of sequencing.

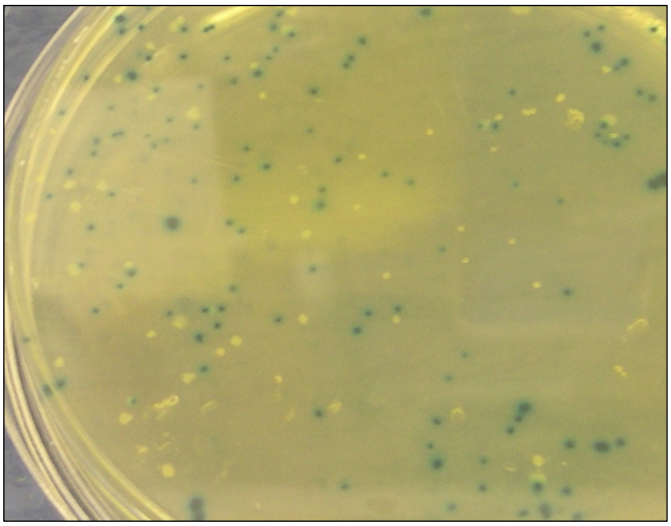


Figure 3.5 Growth of transformants on LB agar containing 5% filter sterilized sucrose and 50µg/ml X-gal. White colonies represent double cross over mutants.

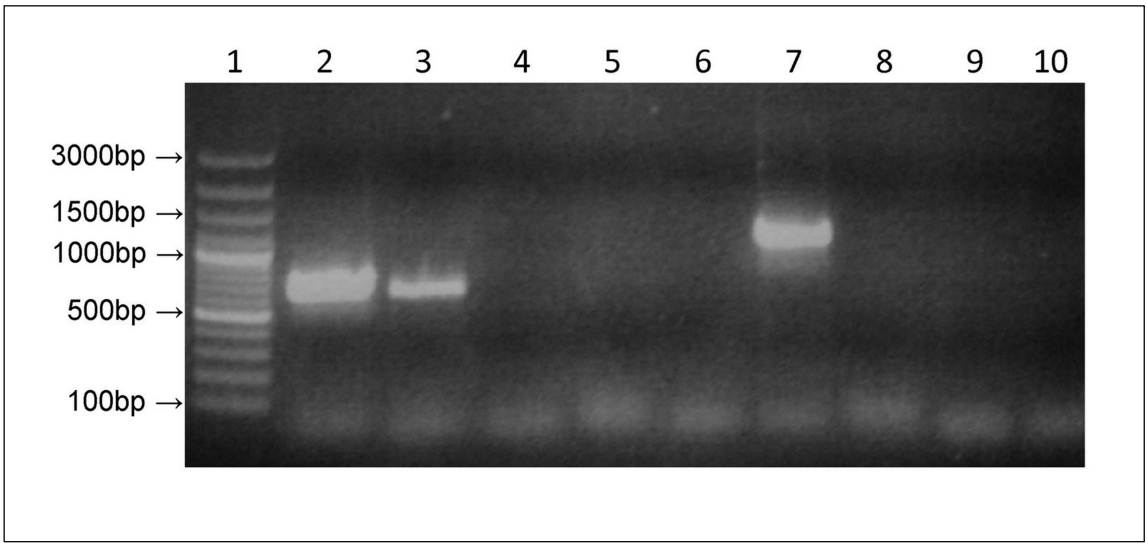


Figure 3.6 Confirmation of DCO knock-out. Lanes 2-4: Wild type *M. smegmatis*; Lanes 5-7: ESX-4 knock-out in *M. smegmatis*; Lanes 8-10: Water blank. PCR amplification to confirm the presence of a DCO knock-out was carried out with Region 4 specific F1 and F2 primers (Table 2.3). Amplification with F1 (f + r) and F2 (f + r) produce products for wild type *M. smegmatis* (lanes 2 and 3), but no products for a knock-out (lanes 5 and 6). Amplification with F1 (f) and F2 (r) produce no product for wild type *M. smegmatis* (lane 4), but does produce a product for the knock-out (lane 7).

CHAPTER FOUR

DISCUSSION

4.1 Mycobacterium Protein Fragment Complementation (M-PFC)

4.1.1 M-PFC Methodology

This study aimed to investigate the protein-protein interactions of the mycobacterial ESAT-6 gene cluster region 4 (ESX-4) in the model organism *Mycobacterium smegmatis* by means of Mycobacterial Protein Fragment Complementation (M-PFC) (Singh et al., 2006). ESX-4 is the most ancestral of the mycobacterial ESAT-6 gene cluster regions of which there are five copies in the genome of *Mycobacterium tuberculosis*. These gene cluster regions encode for a novel secretion system, which is directly involved in pathogenicity (Abdallah et al., 2007).

ESAT-6 region 4 contains 7 genes; these include the genes for secreted T-cell antigens ESAT-6 (esxT) and CFP-10 (esxU), as well as a probable membrane-anchored mycosin protease. Genes encoding a probable conserved integral membrane protein, probable conserved membrane protein ATPase and probable conserved membrane protein are also present. This region also contains one gene that is specific to this region and is a hypothetical alanine and valine rich protein. These genes were divided into 12 constructs and cloned into two M-PFC vectors, pUAB300 and pUAB400, resulting in a total of 24 constructs. All constructs had to be put through numerous steps to ensure the correctness of the final constructs, which included amplification, two rounds of transformation into *Escherichia coli* and purification in order to perform sequencing. The final constructs were co-transformed into the model organism *M. smegmatis* and screened on plates containing kanamycin, hygromycin and trimethoprim.

Growth on plates containing these antibiotics were indicative of potential protein-protein interactions. Many interactions were found throughout the region, including the interaction between ESAT-6 and CFP-10, which are known interacting partners (Renshaw et al., 2002). The significance of these interactions is discussed in section 4.1.2.

M-PFC provides the first experimental system for studying protein-protein interactions in mycobacteria, thus overcoming some of the limitations encountered when using the classical yeast two-hybrid (Y2H) system. *Mycobacterium tuberculosis* has an extraordinarily high

G+C content DNA, which is not always well tolerated in an Y2H system. Furthermore, interactions occur in the nucleus in the Y2H system and self activation can also occur, while bacterial proteins might not undergo the appropriate posttranslational modification. M-PFC aids in overcoming these problems, as well as providing the potential to study protein associations in the mycobacterial membrane (Singh et al., 2006).

This procedure also has the advantage of being dynamic, since previous steps can be repeated for individual constructs, should problems occur in later steps. The ability to include sequencing steps at various points in the procedure help to ensure the correctness of the final constructs.

The M-PFC system seems to be able to identify true interactions. When incubated for an additional week, no additional interacting partners could be identified, indicating that spontaneous growth will not occur for all co-transformations, if incubated for a longer period of time. A trimethoprim concentration of 15µg/ml is also sufficient for screening, with no growth occurring at a concentration of 25µg/ml, which is lower than the original recommended concentration of 50µg/ml trimethoprim (Singh et al., 2006).

It is, however, important that M-PFC only be considered a fishing experiment, since the protein-protein interactions identified by this method need to be verified by additional experiments. Possible procedures could include the Split-Trp protein sensor which is also a mycobacterial two-hybrid system. In this case, should protein interactions occur, growth on a medium containing tryptophan will occur. The Split-Trp method can be applied in both *E. coli* and *M. smegmatis* (O'Hare et al., 2008). More diverse techniques for the confirmation of these reactions could include Biacore assays and pull-down assays. Biacore assays provide a label-free method to analyse interactions in real time and include determining the kinetics and binding strength of interactions (GE Healthcare). A pull-down assay is an *in vitro* technique whereby a purified, tagged protein will serve as the bait, to be "pulled-down" by a protein serving as the prey, should a protein-protein interaction occur (Schechtman et al., 2003).

Future experiments to complete the interactome study of the ESAT-6 region 4 of *M. tuberculosis* by means of M-PFC will include screening for protein interactions between ESAT-6 region 4 members and the proteins of the other ESAT-6 gene cluster regions. This could also be expanded into a screen of the full genome of *M. tuberculosis*, as a means to determine all interactions of this region and to identify novel substrates. Since ESAT-6 region 4 has been shown to be ancestral (Gey van Pittius et al., 2001), discoveries about this specific region will result in a better understanding of the evolution of the ESAT-6 gene cluster regions, as well as provide clues to the original function of such a gene cluster region.

4.1.2 The interactions of ESX-4

Many interactions were identified throughout ESX-4, including the expected interaction between ESAT-6 (Rv3444) and CFP-10 (Rv3445). For the fragments of Rv3444 (Rv3444A and Rv3444B) and Rv3445 (Rv3445A and Rv3445B) very few interactions were found when fused to the pUAB400 vector. Rv3444 and Rv3445 were originally split into two fragments each, because of possible toxic side effects when the entire genes are expressed, but seemed to be more toxic when split. Seven co-transformants, containing at least one split gene each, could not be successfully propagated in liquid culture, which could be indicative of toxicity. Only interactions of the entire Rv3444 and Rv3445 genes will be discussed further.

Not all interactions could be identified in both directions, which could indicate that some refinements are still required in the M-PFC system. It is possible that interference of the fused mDHFR domain could also influence or inhibit the correct folding of the proteins. The fact that there is only one copy of the integrative vector (pUAB400), but possibly multiple copies of the episomal vector (pUAB300) present could also explain why some interactions could not be found in both directions. It is also possible that, even though *M. smegmatis* is a good model organism for *M. tuberculosis*, not all posttranslational modification and protein folding events are able to occur correctly.

These interactions can be used to draw interaction maps of ESX-4. Self-interactions, which are indicative of homomultimerization, and interactions which could be identified in both directions are indicated in Fig 4.1 and a map of all the interactions of ESX-4 is shown in Fig 4.2.

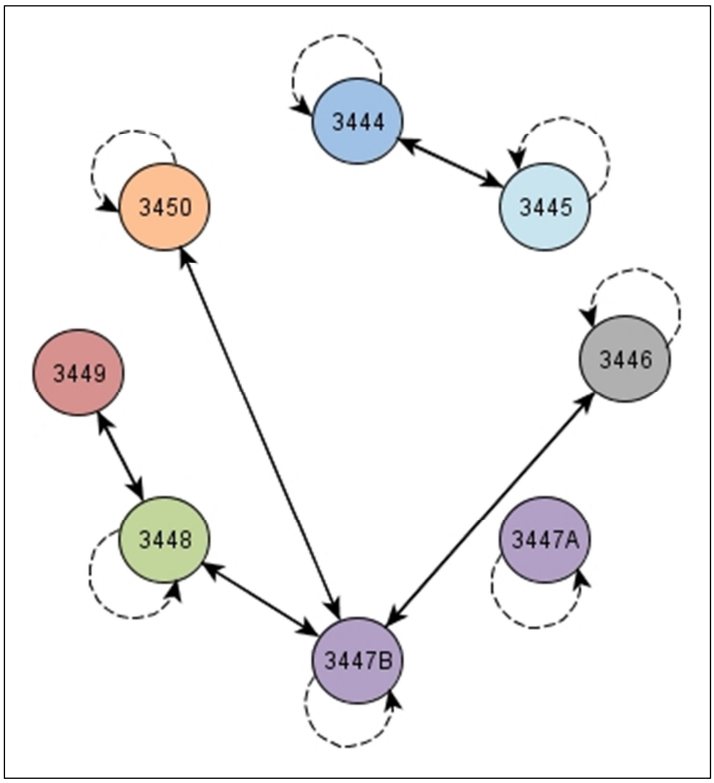


Figure 4.1 Interaction map of the protein interactions of ESX-4. Self interactions are indicated as dashed lines and interactions between proteins are indicated as solid lines. Interactions found in only one direction, as well as interactions of the fragments of Rv3444 (Rv3444A and Rv3444B) and Rv3445 (Rv3445A and Rv3445B) have been excluded.

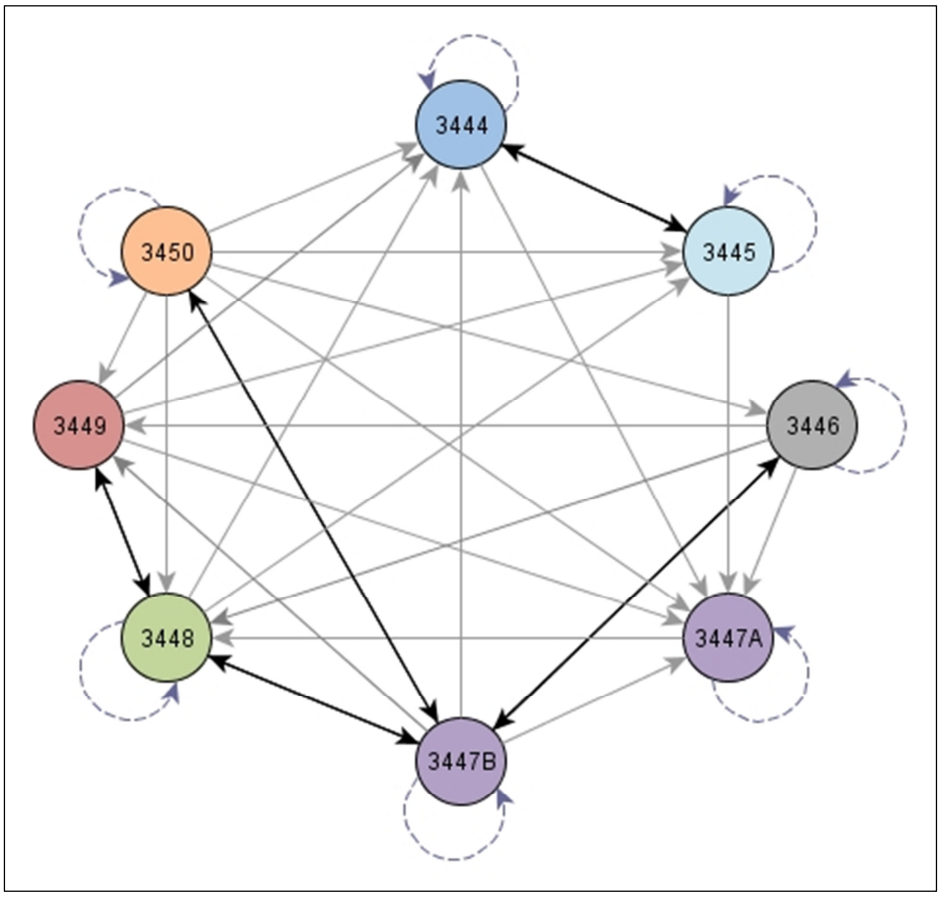


Figure 4.2 Interaction map of the protein interactions of ESX-4. Self interactions are indicated as dashed lines and interactions between proteins are indicated as solid lines. Interactions found in both directions are indicated as solid black lines and interactions found in only one direction are indicated in solid grey lines. Interactions of the fragments of Rv3444 (Rv3444A and Rv3444B) and Rv3445 (Rv3445A and Rv3445B) have been excluded.

These interactions were used to set up a model of the ESX-4 secretion system. The interactions which were found suggest that the components of ESX-4 can assemble into the novel type VII secretion system, through which the ESAT-6/CFP-10 complex is secreted (Fig 4.3). The interactions found throughout the region also suggest that all gene products participate in the formation and function of the ESX secretion machinery (Fig 4.3). Since ESX-4 is ancestral, it represents the most basic components for the functioning of such a secretion system.

Based on the protein-protein interactions identified by M-PFC, the ESX-4 secretion system bears homology to the ESX-1 secretion system. We hypothesize that ESAT-6 (esxT) and CFP-10 (esxU) form a tight complex and are secreted through the pore forming integral membrane protein (Rv3448). The membrane protein ATPase (Rv3447) provides the energy for the active transport system. The region specific alanine and valine rich protein (Rv3446) also shows interaction with the membrane protein ATPase, although the function is not certain. Furthermore, the membrane-anchored mycosin protease (Rv3449) shows interaction with Rv3448, but the exact function is unknown (Fig 4.3).

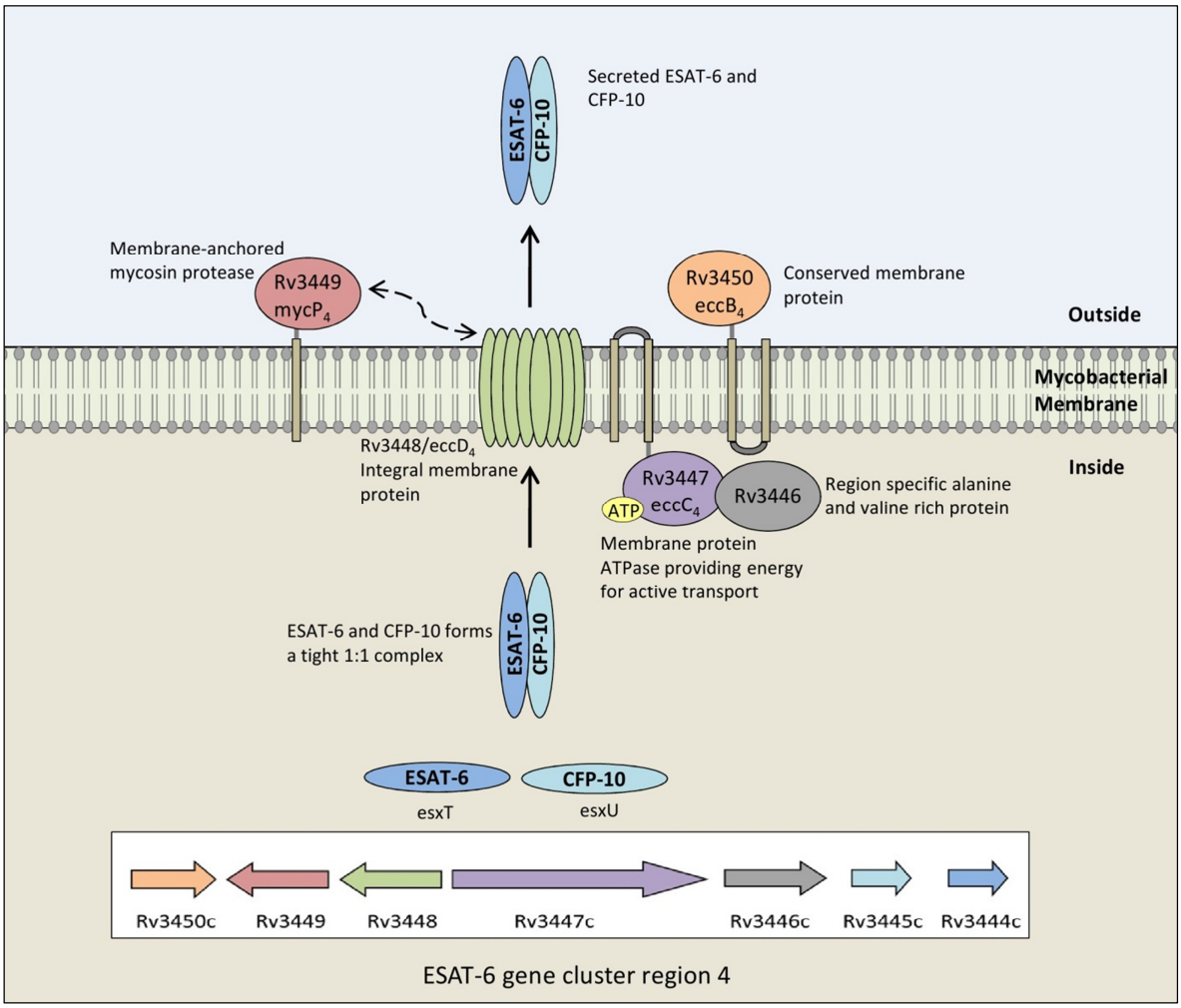


Figure 4.3 Proposed model of the ESX-4 secretion system, containing all the components of ESX-4 and the interactions identified by means of M-PFC. Protein-protein interactions occur throughout the region, with ESAT-6 and CFP-10 forming a tight 1:1 complex, which is then hypothesized to be secreted through the pore-forming integral membrane protein (Rv3448). The membrane protein ATPase (Rv3447) provides the energy for this active transport and shows interaction with the ESX-4 specific protein, Rv3446. A possible interaction between the membrane anchored mycosin protease (Rv3449) and Rv3448 (indicated by the dashed arrow) has also been identified, although the exact function is unknown.

ESX-4 is the ancestral ESX region of *M. tuberculosis*. It also contains the least number of genes of all five regions and is not required for *in vivo* growth of the organism. However,

based on the protein-protein interactions identified by means of M-PFC, the components of ESAT-6 region 4 assemble to form the novel type VII secretion system whereby the ESAT-6/CFP-10 complex is secreted. It would seem that all components form part of this mechanism, although the exact functions of some of the components remain unclear. A comparison of the model of secretion of ESX-4 and the other ESX secretion systems, specifically ESX-1, can provide further insights into which components are directly involved in pathogenicity, since ESX-4 contains only seven genes, of which one is specific to this region.

4.1.3 Comparison of the protein interactions in ESX-4 and ESX-3

The protein-protein interactions of ESX-3 have previously been elucidated by means of M-PFC (Table 4.1). ESX-3 represents a duplication of the ESX gene cluster region. It contains 11 genes and had been shown to be required for mycobactin-mediated iron acquisition (Siegrist et al., 2009).

ESX-4 contains one gene (Rv3446) which is only found in this region and ESX-3 contains five genes not present in ESX-4, namely members of the PE and the PPE families, and genes encoding for an AAA+ class ATPase, ABC transporter family signature and a 2 x amino-terminal transmembrane protein. Thus, the protein-protein interactions of only six genes present in both regions could be compared between ESX-4 and ESX-3. These include ESAT-6, CFP-10, eccC (gene split into two due to size), eccD and eccB, as well as MycP. ESX-3 contains 25 comparable interactions (Table 4.1) and ESX-4 contains 30 comparable interactions (Table 4.2). When compared, 19 protein-protein interactions were found to be identical between the two regions (Table 4.3). This indicates a definite similarity between the protein-protein interactions of ESX-3 and ESX-4, and thus the mechanism of secretion. It is also indicative of a similarity in sensitivity of the M-PFC system when applied to the different ESX regions.

Table 4.1 Comparable protein interactions of ESX-3. Protein interactions are indicated by an 'X', grey blocks represent co-transformations that could not be propagated in liquid culture for ESX-4. (Information kindly supplied by M Newton-Foot)

		PUAB300 Construct						
		esxH Rv0288	*esxG* Rv0287	*eccC$_3$A* Rv0284A	*eccC$_3$B* Rv0284B	*eccD$_3$* Rv0290	*mycP$_3$* Rv0291	*eccB$_3$* Rv0283
pUAB400 Construct	*esxH* Rv0288	X	X	X	X			
	esxG Rv0287	X	X	X				
	eccC$_3$A Rv0284A		X	X	X	X		X
	eccC$_3$B Rv0284B		X	X		X		
	eccD$_3$ Rv0290				X	X		
	mycP$_3$ Rv0291	X	X	X	X	X		
	eccB$_3$ Rv0283			X		X		X

Table 4.2 Comparable protein interactions of ESX-4. Protein interactions are indicated by an 'X', grey blocks represent co-transformations that could not be propagated in liquid culture for ESX-4.

		PUAB300 Construct						
		esxT Rv3444	*esxU* Rv3445	*eccC$_4$A* Rv3447A	*eccC$_4$B* Rv3447B	*eccD$_4$* Rv3448	*mycP$_4$* Rv3449	*eccB$_4$* Rv3450
pUAB400 Construct	*esxT* Rv3444	X	X	X				
	esxU Rv3445	X	X					
	eccC$_4$A Rv3447A		X	X		X		
	eccC$_4$B Rv3447B	X		X	X	X	X	X
	eccD$_4$ Rv3448	X	X		X	X	X	
	mycP$_4$ Rv3449	X	X	X		X		
	eccB$_4$ Rv3450	X	X	X	X	X	X	X

Table 4.3 Overlay of ESX-3/ESX-4 protein interactions, as identified by M-PFC. Interactions occurring in both ESX-3 and ESX-4 are indicated by a bold '**X**', grey blocks represent co-transformations that could not be propagated in liquid culture for ESX-4. Pink shading indicates self-interaction and green shading indicates growth for ESX-4 genes in both vector combinations.

		PUAB300 Construct						
pUAB400 Construct		ESAT-6	CFP-10	*eccC-A*	*eccC-B*	*eccD*	*mycP*	*eccB*
	ESAT-6	**X**	**X**	**X**				
	CFP-10	**X**	**X**					
	eccC-A		**X**	**X**		**X**		
	eccC-B			**X**		**X**		
	eccD				**X**	**X**		
	mycP	**X**	**X**	**X**		**X**		
	eccB			**X**		**X**		**X**

4.2 The construction of genetic knock-outs of ESAT-6 Region 4 in *Mycobacterium smegmatis*

4.2.1 Homologous Recombination Methodology

A further aim of the investigation was to create a genetic knock-out of ESX-4 in *M. smegmatis* by means of homologous recombination techniques (Parish et al., 1999). Homologous recombination proved to be an effective technique in generating targeted genetic knock-outs, while including selectable and counter-selectable markers to ease in the identification of single cross over (SCO) and double cross over (DCO) events. Sensitivity of these markers might be improved by increasing the concentration, since a concentration of 5% sucrose still allows for the growth of many blue colonies in the DCO selection step.

4.2.2 The construction of a genetic knock-out of ESX-4 in *M. smegmatis*

A targeted genetic knock-out of ESX-4 in *M. smegmatis* was constructed by means of homologous recombination and confirmed by means of PCR screening and sequencing. No differences could be observed between the knock-out and wild type strain in terms of growth rate in liquid media, colony morphology on solid media or by microscopy after ZN staining.

The knock-out strain will also be utilised in further comparative studies, which will enable us to distinguish phenotypic differences between the knock-out and wild type strains, as well as to identify variations in their proteomes, secretomes and metabolomes.

CHAPTER FIVE

CONCLUSION AND FUTURE DIRECTIONS

5.1 Mycobacterium Protein Fragment Complementation (M-PFC)

M-PFC is a dynamic mycobacterial two-hybrid system, which provides many advantages for identifying protein-protein interactions in mycobacteria, whilst overcoming the limitations of the classical yeast two-hybrid system. However, it is important to regard M-PFC as a fishing experiment and it is essential that further experimental approaches are applied in order to confirm these interactions.

The M-PFC system will also benefit from refinements, such as acquiring a dam(-) *E. coli* strain with blue/white selection for initial transformation steps which would increase the likelihood of the desired construct being present when colonies are picked and could, in turn, increase chances of finding the construct in the first round of PCR screening.

The protein-protein interactions identified in this study shows that a wide range of interactions occurs within ESAT-6 region 4. It also indicates that the components of ESX-4 assemble to form the novel type VII secretion system whereby the ESAT-6/CFP-10 complex is secreted. It is essential that these findings be confirmed by other techniques in order to set up a robust model of the ESX-4 secretion system based only on confirmed interactions. A comparison of the model of secretion of ESX-4 and the other ESX secretion systems, specifically ESX-1, can provide further insights into which components are directly involved in pathogenicity, since ESX-4 contains only seven genes, of which one is region specific.

The identification of interactions between ESAT-6 region 4 and the other ESAT-6 gene cluster regions, as well as a screen of the full genome of *M. tuberculosis* could lead to discoveries about this region and will result in a better understanding of the evolution of the ESAT-6 gene cluster regions, as well as provide clues to the original function of these a gene cluster regions.

5.2 The construction of genetic knock-outs of ESAT-6 Region 4 in *Mycobacterium smegmatis*

Homologous recombination is an effective method for the construction of targeted genetic knock-outs and a knock-out of ESX-4 in *M. smegmatis* could be obtained. This also indicated that ESX-4 is not essential for the *in vivo* growth of *M. smegmatis*. The knock-out strain also showed no difference in growth rate, colony morphology or by microscopy after ZN staining, when compared to the wild type strain. Future metabolomic, secretomic and proteomic approaches will provide more data on the function of ESX-4.

A future objective would be to construct a double knock-out of ESX-3 and ESX-4 in *M. smegmatis*, whereby only ESX-1 remains present in the genome. This will allow for various functional studies to be carried out on a *M. smegmatis* strain containing only ESX-1. Individual genes of ESX-1 could then be targeted for knock-out in order to study the function of each component of the system. Finally, all three ESX regions of *M. smegmatis* could be knocked out and, if a viable strain is achieved, further metabolomic, secretomic and proteomic studies can be carried out.

In addition, if functional differences are discovered between knock-out strains and the wild type strain, the specific region must be reintroduced into the genome by means of knock-in techniques in order to restore the original function and satisfy molecular Koch's postulate. The two ESX regions absent from *M. smegmatis*, but present in *M. tuberculosis*, namely ESX-2 and ESX-5 could also be knocked into *M. smegmatis* and various functional studies carried out.

A greater understanding of these immunopathogenically important secretion systems may provide valuable clues as to how to interfere with these systems in the pathogen, potentially leading to advances in the treatment and prevention of tuberculosis disease.

CHAPTER SIX

ADDENDUM

6.1 Addendum A

Additional photographic representation of agarose gels used in Mycobacterial Protein Fragment Complementation

PCR Screening

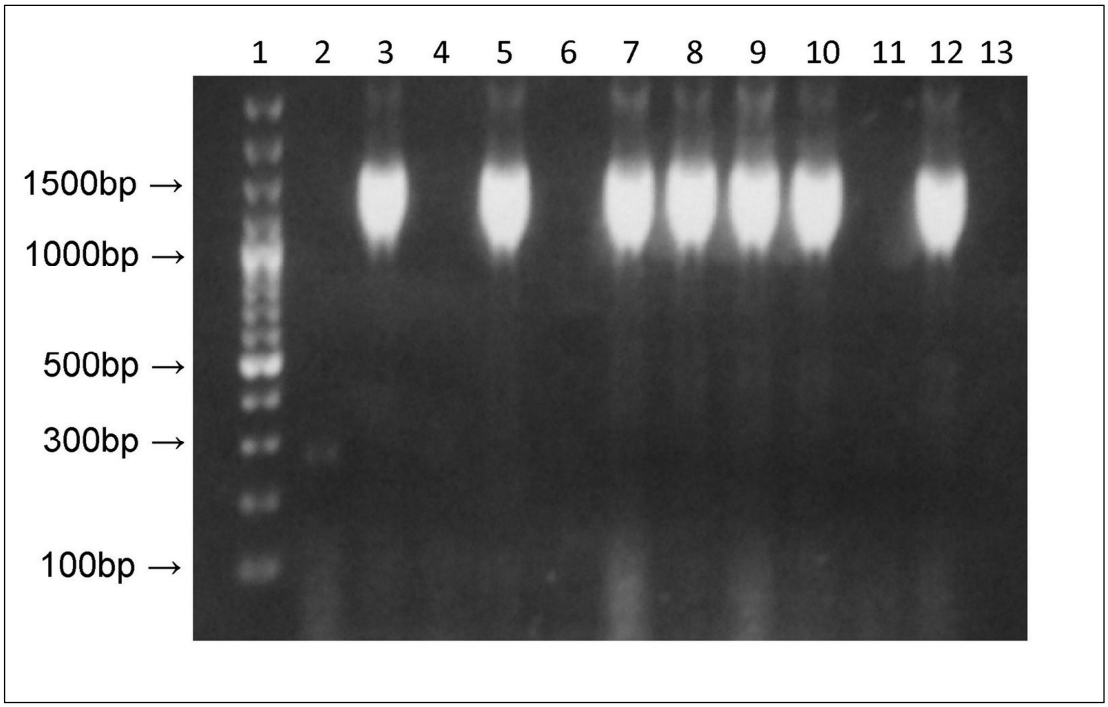


Figure A.1 PCR screening of Rv3449 to confirm presence of insert in pUAB300, as well as orientation. A specific pUAB300 sequencing primer served as the forward primer, while the original reverse primer (Table 2.2). The use of these primers results in a PCR amplification product ± 80bp larger than the original amplification product, if the insert is present and in the correct orientation. Should the insert be present, but in the incorrect orientation, a PCR amplification product of ± 100bp will be produced. In the case of Rv3449 (original product size of ± 1400bp) a product of ± 1500bp is produced in lanes 3,5,7,8,9,10 and 12, indicating the presence of the insert in the correct orientation. All pUAB300 ligated inserts were screened in this manner. pUAB400 inserts were screened in a similar manner, but a pUAB400 specific sequencing primer served as forward primer.

6.2 Addendum B

Additional photographic representation of agarose gels used in knock-out construction

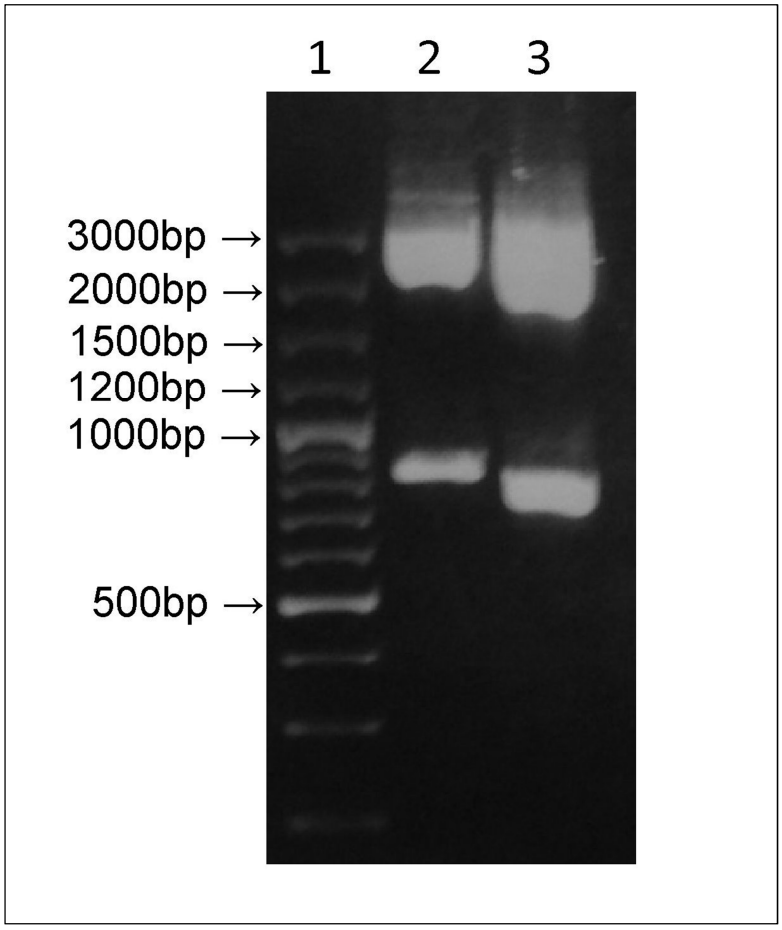


Figure B.1 Digest of the Upstream and Downstream regions of Region 4. The Upstream region was digested out of pGEM-T Easy with the restriction enzymes *Sal*I and *Sfu*I (Lane 2). The Downstream region was digested out of pGEM-T Easy with the restriction enzymes *Sfu*I and *Hind*lII (Lane 3).

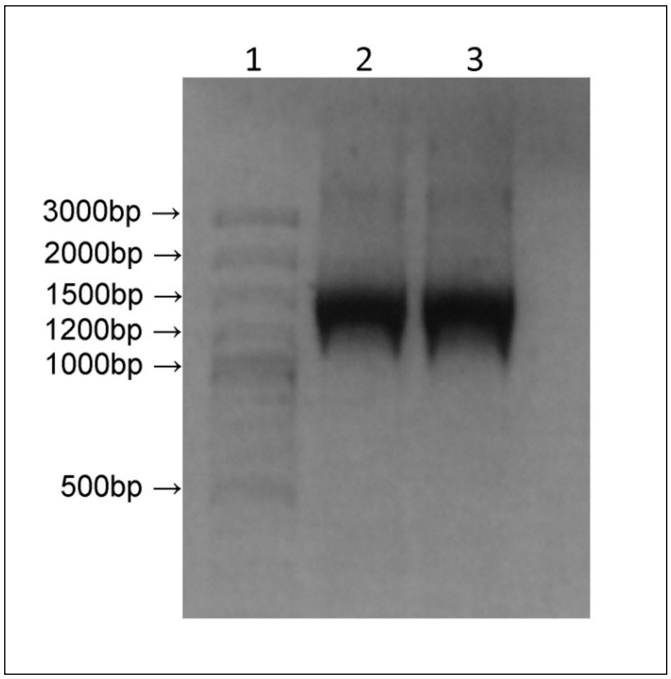


Figure B.2 Confirmation of disruption element. Lanes 2 and 3 indicate two samples potentially containing the disruption element. PCR amplification with original Region-4 Upstream forward primer and Downstream reverse primer (Table 2.3) results in a PCR product of ± 1400bp, confirming the presence of the disruption element.

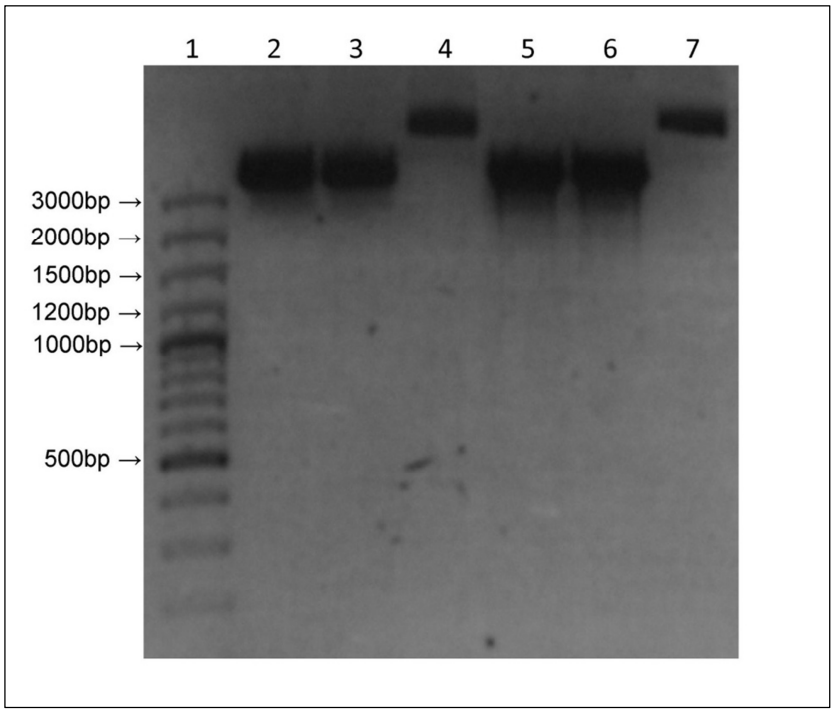


Figure B.3 Digest of p2NIL:ΔR4 with *Pac*I. Two p2NIL mutants containing the disruption element were digested with *Pac*I. Lanes 2 and 3 contain the cut plasmid of the first sample and lanes 5 and 6 contain the cut plasmid of the second sample. Lanes 4 and 7 contain uncut plasmid.

6.3 Addendum C

6.3.1 Preparation of Electro competent *E. coli*

6.3.1.1 *E. coli* K12 ER2925

For the first round of M-PFC *E. coli* transformations the dam negative (-) strain K12 ER2925 was utilised. It was essential to use a strain with dam negative properties in order to prevent methylation of the restriction site.

Competent cells were inoculated into LB media containing 25µg/ml chloramphenicol and grown overnight at 37°C with shaking until an OD_{600} of between 0.6 and 0.8 was reached. All subsequent steps were carried out under icy cold conditions.

6.3.1.2 *E. coli* JM109

All *E.coli* transformations for knock-out construction were carried out in *E. coli* strain JM109, which enables blue/white colony selection in the presence of X-gal.

Competent cells were inoculated into LB media containing 10µg/ml tetracycline and grown overnight at 37°C with shaking until an OD_{600} of between 0.6 and 0.8 was reached. All subsequent steps were carried out under icy cold conditions.

6.3.1.3 General instructions

Cells were harvested in a GSA rotor for 10 minutes at 5000 rpm and re-suspended in a volume of ice-cold 10% glycerol equal to the original culture volume and washed twice at 5000 rpm. All cells were then pooled in a 50ml tube and re-suspended in 10% glycerol. The cells were washed once in a SS34 rotor for 15 minutes at 3000 rpm and re-suspended in 10% glycerol, using a volume of 2ml 10% glycerol per initial culture.

Aliquots of 100µl were transferred to sterile 1.5ml tubes and frozen in a liquid nitrogen bath. Cells were stored at -80°C.

6.3.2 Preparation of Electro competent *M. smegmatis*

M. smegmatis mc^2155 was used for all mycobacterial transformations. 100µl electro competent *M. smegmatis* was inoculated into 10ml 7H9 containing either Kanamycin (25µg/ml) for *M. smegmatis* containing the pUAB400 integrated plasmid (M-PFC), or lacking antibiotic, in the case of wild type *M. smegmatis*. The 10ml starter cultures were incubated overnight at 37°C with shaking.

One millilitre *M. smegmatis* from the 10ml starter culture was inoculated into 100ml 7H9 containing either Kanamycin (25µg/ml) for *M. smegmatis* containing the pUAB400 integrated plasmid (M-PFC), or lacking antibiotic, in the case of wild type *M. smegmatis*. Cultures were grown overnight at 37°C with shaking until an OD_{600} of between 0.5 and 1.0 was reached. All subsequent steps were carried out under icy cold conditions.

The culture was transferred to 50ml tubes and the cells harvested in a SS34 rotor at 4000 rpm for 15 minutes. Cells were re-suspended in 10% glycerol and washed twice at 4000 rpm for 15 minutes. Cells were pooled in one 50ml tube and washed once at 4000 rpm for 10 minutes. Cells were re suspended in 2ml 10% glycerol. Electro competent cells could be used immediately, or frozen in a liquid nitrogen bath and stored at -80°C.

6.4 Addendum D

Examples of M-PFC screening

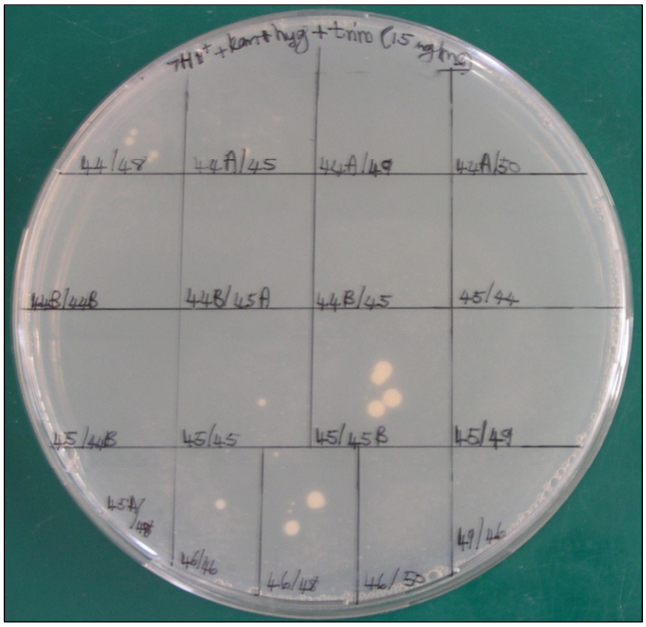


Figure D.1 Example of M-PFC screening on plates containing kanamycin, hygromycin and trimethoprim (15µg/ml).

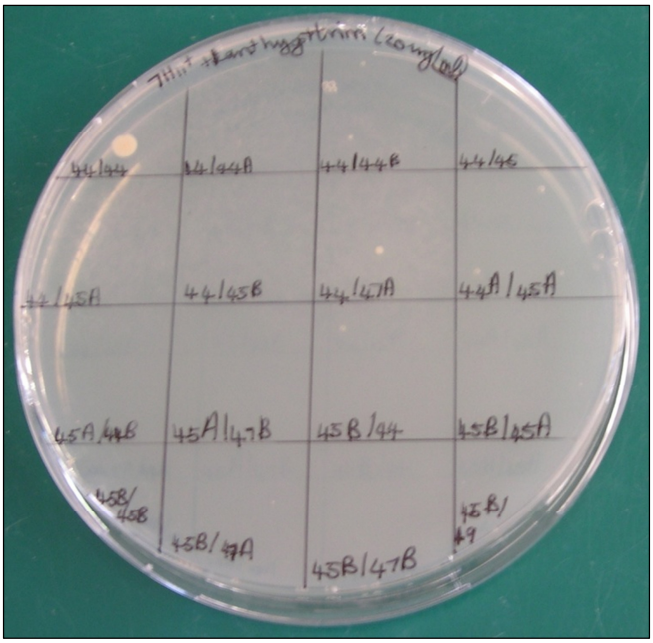


Figure D.1 Example of M-PFC screening on plates containing kanamycin, hygromycin and trimethoprim (20µg/ml).

CHAPTER SEVEN

LIST OF REFERENCES

Bibliography

Abdallah, A., Gey van Pittius, N, DiGiuseppe Champion, P, Cox, J, Luirink, J, Vandenbroucke-Grauls, C, Appelmelk, B and Bitter, W. (2007). Type VII secretion - mycobacteria show the way. *Nature Reviews - Microbiology*, **5**, 883-891.

Aldovini, A., Husson, R and Young, R. (1993). The uraA locus and homologous recombination in *Mycobacterium bovis* BCG. *Journal of Bacteriology*, **175**, 7282-7289.

Arnold, K., Bordoli, L, Kopp, J, and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195-201.

Azad, A., Sirakova, T, Rogers, L and Kolattukudy, P. (1996). Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG producesa mutant that lacks mycosides. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 4787-4792.

Azad, A., Sirakova, T, Fernandes, N and Kolattukudy, P. (1997). Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic Mycobacteria. *The Journal of Biological Chemistry*, **272**, 16741-16745.

Balasubramanian, V., Pavelka, M, Bardarov, S, Martin, J, Weisbrod, T, McAdam, R, Bloom, B and Jacobs, W. (1996). Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *Journal of Bacteriology*, **178**, 273-279.

Bitter, W., Houben, E, Bottai, D, Brodin, P, Brown, E, Cox, J, Derbyshire, K, Fortune, S, Gao, L, Liu, J, Gey van Pittius, N, Pym, A, Rubin, E, Sherman, D, Cole, S and Brosch, R. (2009).

Systematic genetic nomenclature for Type VII secretion systems. *PLos Pathogens*, **5**, e1000507.

Brodin, P., de Jonge, M, Majlessi, L, Leclerc, C, Nilges, M, Cole, S and Brosch, R. (2005). Functional analysis of ESAT-6, the dominant T-cell antigen of *Mycobacterium tuberculosis*, reveals key residues involved in secretion, complex-formation, virulence and immunogenicity. *The Journal of Biological Chemistry*, **280**, 33953 - 33959.

Brosch, R., Gordon, S, Marmiesse, M, Brodin, P, Buchrieser, C, Eiglmeier, K, Garnier, T, Gutierrez, C, Hewinson, G, Kremer, K, Parsons, L, Pym, A, Samper, S, van Soolingen, D and Cole, S. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proceedings of the National Academy of Sciences of the United States of America* , **99**, 3684-3689.

Cole, S., Brosch, R, Parkhill, J, Garnier, T, Churcher, C, Harris, D, Gordon, S, Eiglmeier, K, Gas, S, Barry, I.C, Tekaia, F, Badcock, K, Basham, D, Brown, D, Chillingworth, T, Connor, R, Davies, R, Devlin, K, Feltwell, T, Gentles, S, Hamlin, N, Holroyd, S, Hornby, T, Jagels, K, Krogh, A, McLean, J, Moule, S, Murphy, L, Oliver, K, Osborne, J, Quail, M, Rajandream, M, Rogers, J, Rutter, S, Seeger, K, Skelton, J, Squares, R, Squares, S, Sulston, J, Taylor, K, Whitehead, S and Barrel, B. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537-544.

Cole, S. (2002). Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology*, **148**, 2919-2928.

Converse, S. and Cox, J. (2005). A protein secretion pathway critical for Mycobacterial tuberculosis virulence is conserved and functional in *Mycobacterial smegmatis*. *Journal of Bacteriology*, **187**, 1238-1245.

Guex, N. and Peitsch, M. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* **18**, 2714-2723.

Gey van Pittius, N., Gamieldien, J, Hide, W, Brown, G, Siezen, R and Beyers, A. (2001). The ESAT-6 gene cluster of *Mycobacterial tuberculosis* and other high G+C gram-positive bacteria. *Genome Biology*, **2(10)**, 0044.1-0044.18.

Hershkovitz, I., Donoghue, H, Minnikin, D, Besra, G, Lee, O, Gernaey, A, Galili, E, Eshed, V, Greenblatt, C, Lemma, E, Kahila Bar-Gal, G and Spigelman, M. (2008). Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS one*, **3**.

Herzog, H. (1998). History of Tuberculosis. *Respiration*, **65**, 5-15.

Hsu, T., Hingley-Wilson, S, Chen, B, Chen, M, Dai, A, Morin, P, Marks, C, Padiyar, J, Goulding, C, Gingery, M, Eisenberg, D, Russel, R, Derrick, S, Collins, F, Morris, S, King, C and Jacobs, W. (2003) The primary mechanism of attenuation of bacillus Calmette-Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12420-12425.

http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=gms

http://genolist.pasteur.fr/Tuberculist

http://www.biacore.com/lifesciences/introduction/index.html

Husson, R., James, B and Young, R. (1990). Gene replacement and expression of foreign DNA in Mycobacteria. *Journal of Bacteriology*, **172**, 519-524.

Jerant, A., Bannon, M and Rittenhouse, S. (2000) Identification and management of Tuberculosis. *American Family Physician*, **61**, 2667-2678.

Kalpana, G., Bloom, B and Jacobs, W. (1991). Insertional mutagenesis and illegitimate recombination in Mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 5433-5437.

Machowski, E., Dawes, S and Mizrahi, V. (2005). TB tools to tell the tale - molecular genetic methods for mycobacterial research. *The International Journal of Biochemistry and Cell Biology*, **37**, 54-68.

McAdam, R., Quan, S, Smith, D, Bardarov, S, Betts, J, Cook, F, Hooker, E, Lewis, A, Woollard, P, Everett, M, Lukey, P, Bancroft, G, Jacobs, W and Duncan, K. (2002). Characterization of a *Mycobacterium tuberculosis* H37Rv transposon library reveals insertions in 351 ORFs and mutants with altered virulence. *Microbiology*, **148**, 2975-2986.

McCarthy, O. (2001). The key to the sanatoria. *Journal of the Royal Society of Medicine*, **94**, 413-417.

O'Hare, H., Juillerat, A, Dianišková, P and Johnsson, K. (2008). A split-protein sensor for studying protein-protein interaction in mycobacteria. *Journal of Microbiological Methods*, doi:10.1016/j.mimet.2008.02.008.

Parish, T., Gordhan, B, McAdam, R, Duncan, K, Mizrahi, V and Stoker, N. (1999). Production of mutants in amino acid biosynthesis genes of *Mycobacterium tuberculosis* by homologous recombination. *Microbiology*, **145**, 3497-3503.

Parish, T. and Stoker, N. (2000). Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology*, **146**, 1969-1975.

Pavelka, M. and Jacobs, W. (1999). Comparison of the construction of unmarked deletion mutations in *Mycobacterium smegmatis*, *Mycobacterium bovis* Bacillus Calmette-Guèrin, and

Mycobacterium tuberculosis H37Rv by allelic exchange. *Journal of Bacteriology*, **181**, 4780-4789.

Pelicic, V., Jackson, M, Reyrat, J, Jacobs, W, Gicquel, B and Guilhot, C. (1997). Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 10955-10960.

Prescott, L., Harley, J and Klein, D. (2002). *Microbiology 5th edition*. McGraw-Hill Companies Inc, New York.

Pym, A., Brodin, P, Brosch, R, Huerre, M and Cole, S. (2002). Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterial bovis* BCG and *Mycobacterium microti*. *Molecular Microbiology*, **46**, 709-717.

Renshaw, P., Panagiotidou, P, Whelan, A, Gordon, S, Hewinson, R, Williamson, R and Carr, M. (2002). Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6-CFP-10 complex. *The Journal of Biological Chemistry*, **277**, 21598-21603.

Renshaw, P., Lightbody, K, Veverka, V, Muskett, F, Kelly, G, Frenkiel, T, Gordon, S, Hewinson, R, Burke, B, Norman, J, Williamson, R and Carr, M. (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *The EMBO Journal*, **24**, 2491-2498.

Reyrat, J., Berthet, F and Gicquel, B. (1995). The urease locus of *Mycobacterium tuberculosis* and its utilization for the demonstration of allelic exchange in *Mycobacterium bovis* bacillus Calmette-Guèrin. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 8768-8772.

Sani, M., Houben, E, Geurtsen, J, Pierson, J, de Punder, K, van Zon, M, Wever, B, Piersma, S, Jiménez, C, Daffé, M, Appelmelk, B, Bitter, W, van der Wel, N and Peters, P. (2010). Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathogens*, **6**, e1000794.

Sakula, A. (1983). Carlo Forlanini, inventor of artificial pneumothorax for treatment of pulmonary tuberculosis. *Thorax*, **38**, 326-332.

Schechtman, D., Mochly-Rosen, D and Ron, D. (2003). *Methods in Molecular Biology*, **233**, 325-350.

Schwede, T., Kopp, J, Guex, N, and Peitsch, M. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, **31**, 3381-3385.

Siegrist, M., Unnikrishnan, M, McConnell, M, Borowsky, M, Cheng, T, Siddiqi, N, Fortune, S, Moody, D and Rubin, E. (2009). Mycobacterial ESX-3 is required for mycobactin-mediated iron acquisition. *PNAS*, doi/10.1073/pnas.0900589106.

Singh, A., Mai, D, Kumar, A and Steyn, A. (2006). Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11346-11351.

Smith, I. (2003). *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. *Clinical Microbiology Reviews*, **16**, 463-496.

Snapper, S., Melton, R, Mustafa, S, Keiser, T and Jacobs Jr, W. (1990). Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Molecular Microbiology*, **4**, 1911.

WHO (2009) http://www.who.int/tb/publications/2009/factsheet_tb_2009update_dec09.pdf.

Zink, A., Haas, C, Reischl, U, Szeimies, U and Nerlich, A. (2001). Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *Journal of Medical Microbiology*, **50**, 355-366.